

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO TECNOLÓGICO  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**DETECÇÃO DE FALHAS DE PROCESSOS INDUSTRIAIS EM  
MÚLTIPLOS PONTOS DE OPERAÇÃO VIA ANÁLISE EXTERNA  
LINEAR E NÃO LINEAR**

DISSERTAÇÃO DE MESTRADO

JÚLIO CÉSAR OLIVEIRA FARIA

VITÓRIA

2018

UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO TECNOLÓGICO  
DEPARTAMENTO DE ENGENHARIA ELÉTRICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

**DETECÇÃO DE FALHAS DE PROCESSOS INDUSTRIAIS EM  
MÚLTIPLOS PONTOS DE OPERAÇÃO VIA ANÁLISE EXTERNA  
LINEAR E NÃO LINEAR**

JÚLIO CÉSAR OLIVEIRA FARIA

Dissertação de Mestrado apresentada como  
requisito parcial para obtenção do título de  
Mestre em Engenharia Elétrica.

**Orientador:**

Prof. Dr. Celso José Munaro.

VITÓRIA

2018

Dados Internacionais de Catalogação-na-publicação (CIP)  
(Biblioteca Setorial Tecnológica,  
Universidade Federal do Espírito Santo, ES, Brasil)

---

F224d Faria, Júlio César Oliveira, 1988-  
Detecção de falhas de processos industriais em múltiplos  
pontos de operação via análise externa linear e não linear / Júlio  
César Oliveira Faria. – 2018.  
65 f. : il.

Orientador: Celso José Munaro.  
Dissertação (Mestrado em Engenharia Elétrica) –  
Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Localização de falhas (Engenharia). 2. Análise  
multivariada. 3. Modelos lineares (Estatística). 4. Modelos não  
lineares (Estatística). 5. Controle de processo. 6. Análise  
externa. I. Munaro, Celso José. II. Universidade Federal do  
Espírito Santo. Centro Tecnológico. III. Título.

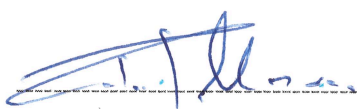
CDU: 621.3

**JÚLIO CÉSAR OLIVEIRA FARIA**

**DETECÇÃO DE FALHAS DE PROCESSOS INDUSTRIAIS EM  
MÚLTIPLOS PONTOS DE OPERAÇÃO VIA ANÁLISE EXTERNA  
LINEAR E NÃO LINEAR**

Dissertação submetida ao programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do Grau de Mestre em Engenharia Elétrica.

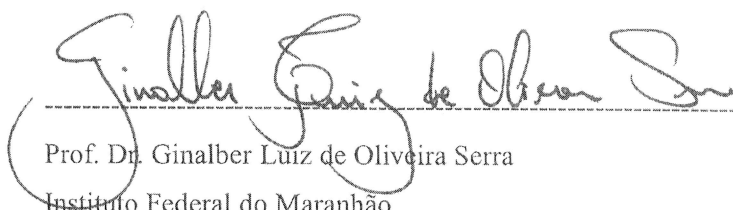
**COMISSÃO EXAMINADORA**



Prof. Dr. Celso José Munaro

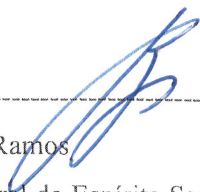
Universidade Federal do Espírito Santo

Orientador



Prof. Dr. Ginalber Luiz de Oliveira Serra

Instituto Federal do Maranhão



Prof. Dr. Rogério Ramos

Universidade Federal do Espírito Santo

## **Agradecimentos**

A minha família pelo incentivo constante ao longo desta jornada, aos colegas da Vale pela oportunidade e suporte, a equipe do projeto de parceria entre Vale e UFES pelas discussões e contribuições essenciais para a elaboração deste trabalho e ao professor Dr. Celso José Munaro pela dedicação, paciência e empenho na orientação desta dissertação.

## Resumo

Os métodos de controle estatístico multivariado de processos com foco em detecção de falhas têm grande potencial de extrair informações relevantes dos dados gerados por instrumentos e sistemas de controle das plantas industriais, que normalmente são subutilizados. Porém, as técnicas tradicionais de controle estatístico multivariado não devem ser utilizadas para processos que operam em múltiplos pontos de operação, devido a incapacidade de distinguir uma falha de uma mudança normal de operação. Essa limitação dificulta o uso destas técnicas em processos reais. Neste contexto, este trabalho apresenta estudos e propõe cinco métodos baseados em Análise Externa não linear e Análise Externa com múltiplos modelos lineares, para detecção de falhas em processos industriais, que naturalmente são não lineares e trabalham em múltiplos pontos de operação. Estes métodos serão aplicados a um simulador *benchmark* da literatura e no monitoramento de um processo real de vibrações de um ventilador de processo de grande porte, usado em um forno de pelletização de minério de ferro. Os resultados mostram que os métodos propostos conseguem distinguir falhas de variações normais de pontos de operação de processos industriais, mantendo o nível de alarmes falsos no valor especificado. Adicionalmente, os resultados mostram que estes métodos têm potencial de detectar falhas automaticamente de forma antecipada, possibilitando ações corretivas que podem diminuir ou até evitar danos a equipamentos de determinado processo, gerando um potencial de ganhos financeiros.

## **Abstract**

The methods of multivariate statistical control of processes focused on fault detection have great potential to extract relevant information from the data generated by instruments and control systems of industrial plants, which are usually underutilized. However, traditional multivariate statistical control techniques should not be used for processes operating at multiple points of operation due to the inability to distinguish a failure from a normal change of operation. This limitation makes it difficult to use these techniques in real processes. In this context, this work presents studies and proposes five methods based on Nonlinear External Analysis and External Analysis with multiple linear models, to detect failures in industrial processes, which are naturally nonlinear and work at multiple points of operation. These methods will be applied to a literature benchmark simulator and to the monitoring of a real vibration process of a large process fan used in an iron ore pelletizing furnace. The results show that the proposed methods can distinguish failures from normal variations of operating points of industrial processes, keeping the level of false alarms in the specified value. Additionally, the results show that these methods have the potential to detect failures automatically in advance, allowing for corrective actions that may reduce or even avoid damages to equipment of a certain process, generating a potential for financial gains.

## Sumário

1	Introdução.....	8
1.1	Objetivos e estrutura da dissertação .....	9
1.2	Contribuições da dissertação e publicações .....	9
2	Revisão da Literatura .....	11
2.1	Controle Estatístico de Processo Multivariado .....	11
2.2	Análise Externa.....	13
2.3	Análise de Componentes Principais ( <i>Principal Component Analysis</i> - PCA).....	15
3	Metodologia proposta.....	18
3.1	Análise Externa com Múltiplos Modelos Lineares - AEML.....	21
3.1.1	Treinamento - AEML .....	22
3.1.2	Monitoramento de falhas - AEML.....	23
3.2	Análise Externa com Múltiplos Modelos Lineares Baseado em RANSAC - AEML-R....	25
3.2.1	Treinamento - AEML-R .....	25
3.2.2	Monitoramento de falhas – AEML-R.....	27
3.3	Análise Externa não linear.....	29
3.3.1	Treinamento – Análise Externa não linear .....	29
3.3.2	Monitoramento de falhas – Análise Externa não linear .....	30
4	Estudos de caso .....	31
4.1	Reator contínuo do tipo tanque agitado ( <i>Continuous Stirred Tank Reactor</i> - CSTR) ..	31
4.2	Monitoramento de vibrações de um ventilador de forno de pelotização.....	41
5	Conclusão .....	53
6	Referências Bibliográficas .....	55



## 1 Introdução

Técnicas de controle estatístico de processos (CEP) vêm sendo largamente utilizadas para o monitoramento de processos industriais, sendo cruciais para analisar a crescente quantidade de dados gerados por instrumentos e equipamentos de forma a contribuir para a melhoria contínua de produtividade e qualidade dos produtos (C. Tong, 2013).

Cartas de controle Shewhart, CUSUM e EWMA são técnicas de controle estatístico univariado consolidadas há décadas, geralmente aplicadas a indicadores críticos do processo, chamados de *Key Performance Indicators* (KPIs).

Técnicas de controle estatístico baseadas em Análise de Componentes Principais (*Principal Component Analysis* - PCA), Mínimos Quadrados Parciais (*Partial Least Squares* - PLS) e suas variantes dinâmicas e recursivas são as metodologias multivariadas mais difundidas, geralmente usadas para monitorar variáveis internas de um processo (C. Tong, 2013). Diferente do CEP univariado, essas técnicas levam em consideração as correlações entre as variáveis e reduzem sua dimensionalidade, de forma a simplificar as análises e desprezar informações não relevantes.

As técnicas de CEP tradicionais não conseguem diferenciar uma mudança de ponto de operação de uma falha de processo, desta forma, elas devem ser usadas para monitorar variáveis com comportamento estacionário, que trabalham em um único ponto de operação (S. J. Zhao, 2004). Esta limitação não dificulta o uso das técnicas univariadas, pois os KPIs de uma planta industrial que são alvos destas análises, geralmente possuem apenas um ponto de operação. Entretanto, as técnicas de CEP multivariado têm como objetivo monitorar variáveis internas de um processo industrial, que normalmente trabalham em múltiplos pontos de operação, devido a demandas distintas de produção, variações nas características dos insumos, entre outros fatores (M. Kano, 2004). Por não conseguir diferenciar falhas de mudanças de ponto de operação, o uso destas técnicas multivariadas em processos reais gera muitos alarmes falsos, criando resistência ao seu uso.

Neste contexto, é essencial o desenvolvimento de técnicas de CEP multivariado que sejam capazes de distinguir variações normais dos pontos de operação e falhas no processo.

A técnica chamada de Análise Externa proposta por Kano (2004) é uma metodologia que se destaca para tratar deste problema. Nesta metodologia, as variáveis do processo

são divididas em dois grupos, variáveis externas e principais, e assume-se que as variáveis principais podem ser explicadas pelas variáveis externas através de um modelo de regressão. Os resíduos resultantes da diferença entre as variáveis principais medidas e estimadas são utilizados para a detecção de falhas e diagnósticos. Desta forma, as variáveis monitoradas passam a ser independentes de variações da região de operação, podendo ser analisadas pelas técnicas tradicionais de CEP multivariado. Entretanto, originalmente esta metodologia utiliza modelo de regressão linear, dificultando o seu uso para processos industriais reais.

### **1.1 Objetivos e estrutura da dissertação**

Esta dissertação tem como objetivo estudar, propor e comparar metodologias baseadas na junção de CEP multivariado e análise externa linear e não linear para detectar falhas em processos industriais que possuem múltiplos pontos de operação.

Este trabalho está organizado da seguinte forma: no capítulo 2 é feita uma revisão da literatura, apresentando os métodos mais difundidos de CEP multivariado usados para detecção de falhas, assim como a Análise Externa. No capítulo 3 são apresentadas cinco metodologias de monitoramento de falhas de processos industriais que possuem múltiplos pontos de operação. No capítulo 4 são apresentados dois estudos de caso, onde as cinco técnicas propostas no capítulo 3 e a análise externa proposta por Kano (2004) são aplicadas a um simulador benchmark da literatura e no monitoramento de um processo real de vibrações de um ventilador de processo de grande porte, usado em um forno de pelotização de minério de ferro da mineradora Vale S.A. Por fim, o capítulo 5 traz as conclusões finais e sugestões de trabalhos futuros.

### **1.2 Contribuições da dissertação e publicações**

Outros trabalhos da literatura já apresentaram técnicas de análise externa não linear em conjunto com técnicas de controle estatístico multivariado com o objetivo de detecção de falhas de processos industriais (Ge, 2008).

Os procedimentos para obter os múltiplos modelos lineares dos dois primeiros métodos propostos no capítulo 3 apresentam originalidade, enquanto que para os outros métodos propostos, técnicas de modelagem não lineares já difundidas na literatura em conjunto com análise externa e PCA foram utilizadas, sendo que o uso dos métodos de

regressão polinomial e *Random Forest* na abordagem de análise externa não foram encontradas na literatura.

Foram geradas as seguintes publicações associadas a esta dissertação: (Faria, J., 2017) e (Assis, A., 2017).

## 2 Revisão da Literatura

Neste capítulo é apresentada a revisão da bibliografia, sendo que a seção 2.1 apresentará as técnicas de controle estatístico multivariado mais estudadas e as técnicas que vêm sendo desenvolvidas para monitorar processos em múltiplos pontos de operação. Na seção 2.2 é detalhado o método de Análise Externa e na seção 2.3 é apresentado o uso de PCA na estrutura de monitoramento de falhas em conjunto com a Análise Externa.

### 2.1 Controle Estatístico de Processo Multivariado

O controle estatístico de processo multivariado é uma das classes de metodologias baseadas em dados históricos utilizadas para detecção de falhas. A Figura 1 mostra uma classificação destes métodos, separando por técnicas qualitativas e quantitativas.

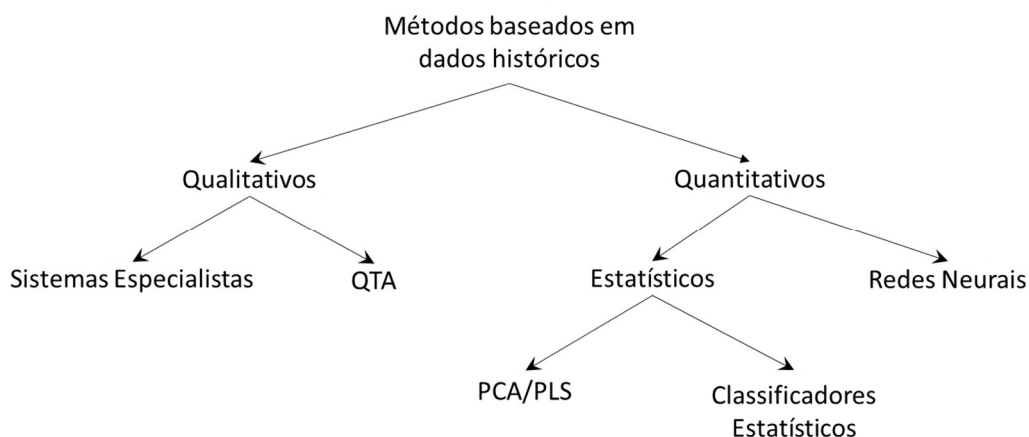


Figura 1 – Classificação de métodos baseados em dados históricos (Venkatasubramanian, 2003)

As técnicas mais difundidas e estudadas de CEP multivariado para monitoramento de processos industriais são PCA e PLS. Essas técnicas têm a capacidade de trabalhar com um volume grande de dados e conseguem reduzir a dimensionalidade dos processos, desprezando informações não relevantes, de forma a simplificar as análises e aumentar a sensibilidade às falhas. Na Figura 2 são mostrados exemplos de artigos que utilizam essas técnicas e suas variantes.

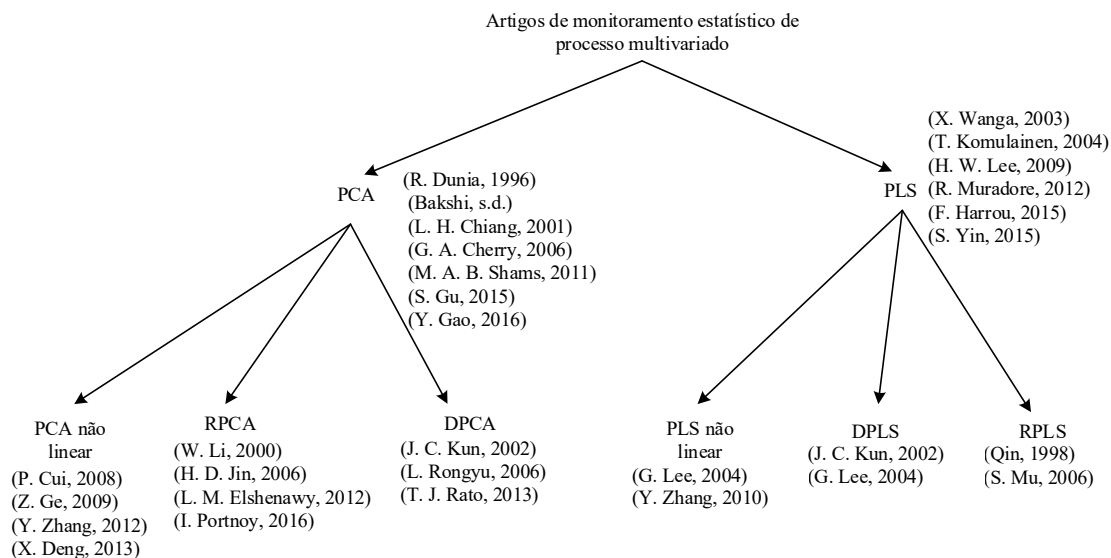


Figura 2 - Artigos relacionados a PCA , PLS e suas variantes usadas como ferramentas de monitoramento de processos

As variações destas metodologias mostradas na Figura 2, foram desenvolvidas para serem aplicadas a processos com características específicas:

- Para processos em que a dinâmica não pode ser desprezada, utiliza-se atrasos das variáveis de processo como novas variáveis. A variação do PCA incorporando a dinâmica é chamado de PCA dinâmico (*Dynamic Principal Component Analysis* - DPCA) ou PLS dinâmico (*Dynamic Partial Least Squares*– DPLS) para o caso de se usar o PLS.
- Para se trabalhar com processos variantes no tempo, é essencial o uso de modelos adaptativos, também chamados de algoritmos recursivos, dando origem as metodologias PCA recursiva (*Recursive Principal Component Analysis* – RPCA) e PLS recursivo (*Recursive Partial Least Squares* – RPLS).
- Para processos não lineares, outras variações foram desenvolvidas, algumas delas chamadas de *Kernel PCA* (KPCA) e *Kernel PLS* (KPLS).

Alguns autores utilizam combinações das variações listadas acima para processos não lineares variantes no tempo (X. Liu, 2009), (M. Ding, 2010), (I. B. Khediri, 2011) e processo não lineares, dinâmicos e variantes no tempo (Y. Zhang, 2012).

Tipicamente, os cálculos das estatística  $T^2$  de Hotelling e estatística  $Q$ , também chamada de soma de erro de predição (*sum of prediction error* - SPE), são usados para

detectar falhas na estrutura de monitoramento de processos baseados em CEP multivariado a partir da comparação com um limiar estatístico (Qin, 2012). Como alternativa uma combinação das duas estatísticas acima pode ser utilizada.

Todas as técnicas citadas acima não devem ser utilizadas para monitorar variáveis que possuem múltiplos pontos de operação, visto que não possuem a capacidade de diferenciar uma mudança normal de ponto de operação de uma falha de processo. Em S. J. Zhao (2004) e G. Yong (2015) é proposto o uso de múltiplos modelos PCA, sendo um modelo treinado para cada região de operação. Porém, esta metodologia requer o treinamento de muitos modelos, tendo em vista que o modelo somente vale em uma estreita faixa de operação, o que também torna o monitoramento mais complexo, visto que é necessária técnicas para identificar qual modelo deverá ser utilizado para detecção de falhas de uma nova amostra. Alternativas para monitoramento de processos com múltiplos pontos de operação podem ser encontradas em J. Yu (2008), S. Tan (2012) e C. Tong (2013).

## 2.2 Análise Externa

Nesta secção é apresentada a metodologia de Análise Externa, proposta por Kano (2004). Os conceitos desta técnica são utilizados nas metodologias propostas neste trabalho, apresentadas no capítulo 3. Considere uma matriz de dados de treinamento  $\mathbf{X} \in \mathbb{R}^{n \times m}$  de um determinado processo, onde  $n$  e  $m$  são o número de amostras e variáveis, respectivamente. As variáveis de  $\mathbf{X}$  podem ser divididas em dois grupos:

$$\mathbf{X} = [\mathbf{H} \ \mathbf{G}] \quad (1)$$

onde  $\mathbf{G} \in \mathbb{R}^{n \times m_G}$  é uma matriz que contém as variáveis externas  $m_G$ , que determinam os pontos de operação do processo e que normalmente são variáveis de entrada, e  $\mathbf{H} \in \mathbb{R}^{n \times m_H}$  é uma matriz que contém as variáveis principais  $m_H$ , que são as variáveis a serem monitoradas e que são afetadas por  $\mathbf{G}$ . A matriz  $\mathbf{H}$  pode ser decomposta em duas partes:

$$\mathbf{H} = \hat{\mathbf{H}} + \mathbf{E} \quad (2)$$

onde  $\hat{\mathbf{H}}$  é a parte de  $\mathbf{H}$  explicada por  $\mathbf{G}$ , e  $\mathbf{E}$  são os resíduos de  $\mathbf{H}$ , que não são explicados por  $\mathbf{G}$ . Para processos com não linearidades desprezíveis, pode ser utilizado um modelo

de regressão linear baseado em mínimos quadrados ordinários (MQO), tendo  $\mathbf{G}$  como entrada e  $\mathbf{H}$  como saída. Desta forma, a matriz de coeficientes de regressão  $\mathbf{C}$  pode ser obtida:

$$\mathbf{C} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{H} \quad (3)$$

A partir de  $\mathbf{C}$ , as variáveis principais estimadas podem ser calculadas:

$$\hat{\mathbf{H}} = \mathbf{G}\mathbf{C} \quad (4)$$

Espera-se que os resíduos  $\mathbf{E}$  sejam independentes dos pontos de operação, tendo em vista que a influência das variáveis que determinam os pontos de operação foi retirada das variáveis principais, sendo possível o uso de técnicas tradicionais de controle estatístico multivariado para detecção de falhas. Os resíduos  $\mathbf{E}$  podem ser calculados da seguinte forma:

$$\mathbf{E} = \mathbf{H} - \mathbf{G}\mathbf{C} \quad (5)$$

Para processos em que as não linearidades não podem ser desprezadas, modificações nesta metodologia precisam ser feitas. Z. Ge (2008) propôs uma metodologia de Análise Externa não linear baseada em SVR (*support vector regression*) e Y. Zheng (2016) propôs o uso de Análise Externa em conjunto com uma variação de PLS, chamada de Projeção Concorrente para Estruturas Latentes (*Concurrent Projection to Latent Structures - CPLS*) proposta por S. J. Qin (2013). Autores como Patan (2008) não mencionam a metodologia de Análise Externa, porém utilizam os mesmos conceitos em suas técnicas para modelos baseados em redes neurais, como mostrado na Figura 3, onde são utilizadas variáveis de entrada de um processo para explicar variáveis de saídas e gerar resíduos, que serão avaliados com foco no monitoramento de falhas.

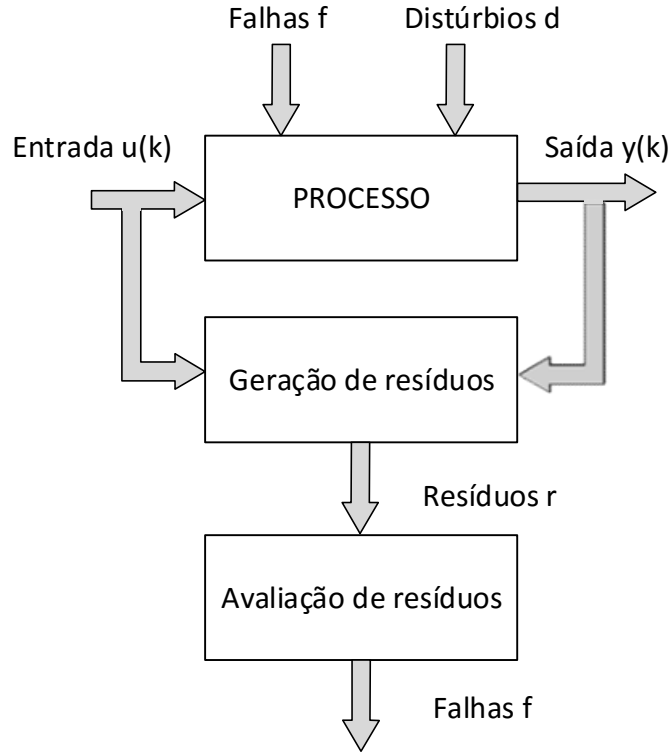


Figura 3 – Monitoramento de falhas baseados na avaliação de resíduos (Patan, 2008)

### 2.3 Análise de Componentes Principais (*Principal Component Analysis - PCA*)

O uso de PCA em uma estrutura de monitoramento de falhas baseada em Análise Externa mapeia as amostras dos resíduos do modelo de Análise Externa  $\mathbf{E}$ , no espaço de componentes principais que são ortogonais e possuem dimensão reduzida. O modelo PCA é obtido a partir de um conjunto de  $n$  amostras sem a presença de falhas, contidas na matriz de resíduos de Análise Externa  $\mathbf{E} \in \Re^{n \times m}$ , calculado a partir da equação (5). É premissa para o uso de PCA que os resíduos sejam sinais temporais aleatórios com distribuição gaussiana.

A matriz de covariância amostral de  $\mathbf{E}$  pode ser calculada da seguinte forma:

$$\mathbf{\Sigma} = \frac{\mathbf{E}^T \mathbf{E}}{n-1} \quad (6)$$

onde  $\mathbf{\Sigma}$  é a matriz de covariância amostral. A partir da decomposição de valores singulares de  $\mathbf{\Sigma}$ :

$$\mathbf{\Sigma} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \quad (7)$$



podemos encontrar a matriz diagonal  $\Lambda$ , que contém os autovalores de  $\Sigma$  em ordem decrescente ( $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$ ), que representam a variância de cada componente e a matriz composta pelos autovetores de  $\Sigma$ , identificada como  $V$ .

Componentes que possuem pouca variância devem ser desprezadas, de forma a diminuir a dimensionalidade do sistema e evitar problemas numéricos, tendo em vista que a matriz contendo os autovalores  $\Lambda$  é invertida no cálculo da estatística  $T^2$ , como mostrado na equação 9. Vários métodos para determinar as componentes que serão desprezadas foram desenvolvidos, como o de porcentagem acumulada de variância (*cumulative percent variance* - CPV), análise paralela e predição da soma do quadrado dos resíduos (*prediction residual sum of squares* - PRESS) (L. H. Chiang, 2001). Neste trabalho utilizaremos o método CPV, que determina a quantidade mínima de componentes  $d$  necessárias para explicar uma porcentagem da variância total e despreza o restante. A variância total retida será de 90%.

Para mapear uma nova amostra de resíduos do modelo de Análise Externa no espaço das componentes principais são utilizadas as  $d$  componentes retidas dos autovetores que estão associadas aos maiores autovalores:

$$\mathbf{t} = \mathbf{e}V_r \quad (8)$$

onde  $\mathbf{e} \in \Re^{1 \times m}$  é uma nova amostra dos resíduos,  $\mathbf{t} \in \Re^{1 \times d}$  representa os valores dos resíduos mapeados nas componentes principais e  $V_r \in \Re^{m \times d}$  é a matriz formada pelos autovetores das  $d$  componentes principais não descartadas.

Para realizar o monitoramento de falhas novas amostras de  $\mathbf{e}$  são projetadas nas componentes principais e a estatística de Hotelling pode ser calculada:

$$T^2 = \mathbf{t}^2 \Lambda_r^{-1} \mathbf{t} = \mathbf{e}V_r \Lambda_r^{-1} V_r^T \mathbf{e}^T \quad (9)$$

onde  $T^2$  é a estatística de Hotelling.

O limite de controle estatístico para a estatística de Hotelling é então calculado:

$$T_\alpha^2 = \frac{d(n^2-1)}{n(n-d)} F_\alpha(d, n-d) \quad (10)$$

onde  $T_\alpha^2$  é o limiar estatístico e  $F_\alpha(d, n - d)$  é a distribuição F, com graus de liberdade  $n$  e  $n - d$  e nível de confiança  $\alpha$ .

A estatística  $\mathbf{Q}$ , também chamada de SPE, também pode ser calculada:

$$\mathbf{r} = (\mathbf{I} - \mathbf{V}_r \mathbf{V}_r^T) \mathbf{e} \quad (11)$$

$$\mathbf{Q} = \mathbf{r}^T \mathbf{r} \quad (12)$$

onde  $\mathbf{I}$  é a matriz identidade de mesma dimensão que  $\mathbf{V}_r \mathbf{V}_r^T$ .

O limite de controle estatístico para a estatística  $\mathbf{Q}$  é calculado usando a distribuição  $\mathbf{Q}$ , que pode ser aproximada pelas seguintes equações:

$$Q_\alpha = \left[ \theta_1 \frac{h_0(1-\alpha)\sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad (13)$$

$$\theta_i = \sum_{j=d+1}^m \sigma_j^{2i} \quad (14)$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \quad (15)$$

onde  $Q_\alpha$  é o limiar estatístico e  $\alpha$  é o nível de confiança.

### 3 Metodologia proposta

Neste capítulo são propostos cinco métodos baseados em Análise Externa e PCA para detectar falhas em processos industriais que operam em múltiplos pontos de operação e que possuem não linearidades. Dois métodos são baseados em Análise Externa com múltiplos modelos lineares e três métodos são baseados em Análise Externa com modelos não lineares.

Estes métodos são divididos em duas etapas:

- Treinamento: etapa *off line* usada para geração dos modelos de análise externa e PCA a partir de amostras do processo em situação de normalidade
- Monitoramento de falhas: etapa *on line* onde novas amostras são disponibilizadas e a detecção de falha pode ocorrer. São utilizados somente dados históricos neste trabalho, sendo simulado o monitoramento de processo *on line*

Os métodos com múltiplos modelos lineares utilizam modelos de regressão estáticos, com a seguinte estrutura:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{m1} \\ 1 & x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{mn} \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (16)$$

onde  $y_i$  é a amostra  $i$  da variável de saída,  $x_{ij}$  é a amostra  $i$  da variável de entrada  $j$ ,  $b_j$  é o coeficiente de regressão da variável  $j$  e  $e_i$  é o erro de modelagem da amostra  $y_i$ .

A estrutura usada não possui parte auto-regressiva, tendo em vista que o objetivo da análise externa é somente retirar a influência das variáveis externas sobre as variáveis principais.

Para justificar o uso destes métodos, é utilizado um exemplo de válvula de controle, comum em todo tipo de indústria, simulada pelas equações abaixo:

$$x(k) = \frac{0,0616q^{-1} + 0,0543q^{-2}}{(1 - 1,5714q^{-1} + 0,687q^{-2})} u(k) \quad (17)$$

$$y(k) = \frac{x(k)}{\sqrt{0,1 + 0,9x(k)^2}} \quad (18)$$

apresentadas por Al-Duwaish e Naem (2001), sendo  $u(k)$  a pressão de controle,  $x(k)$  a posição da haste da válvula e  $y(k)$  o fluxo através da válvula. A Figura 4 mostra graficamente a relação entre  $x(k)$  e  $y(k)$ .

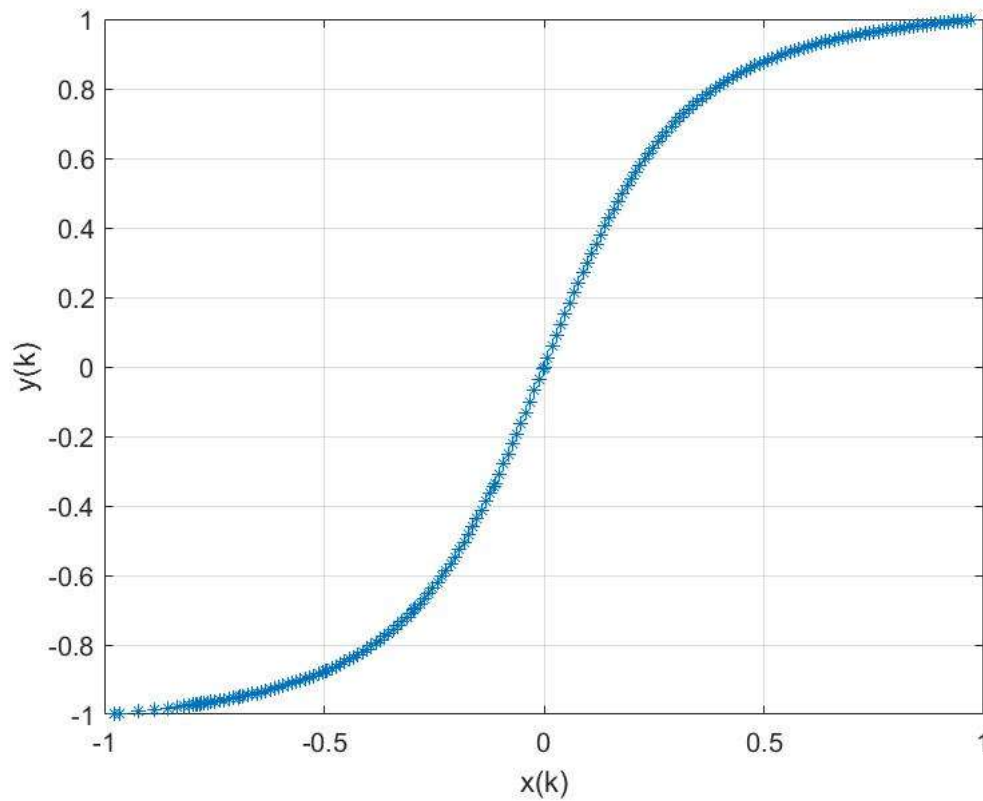


Figura 4 - Válvula de controle (posição x fluxo)

A partir da equação (17) e Figura 4, observa-se uma relação não linear entre a posição e fluxo da válvula, o que é normal em processos industriais reais.

A Figura 5a identifica graficamente um modelo de regressão linear baseado em MQO, tendo como entrada  $x(k)$  e saída  $y(k)$ , a Figura 5b mostra um exemplo de múltiplos modelos lineares e a Figura 5c mostra um modelo não linear polinomial de sexta ordem, enquanto que as Figura 5d-f mostram os respectivos resíduos dos modelos em função de  $x(k)$ .

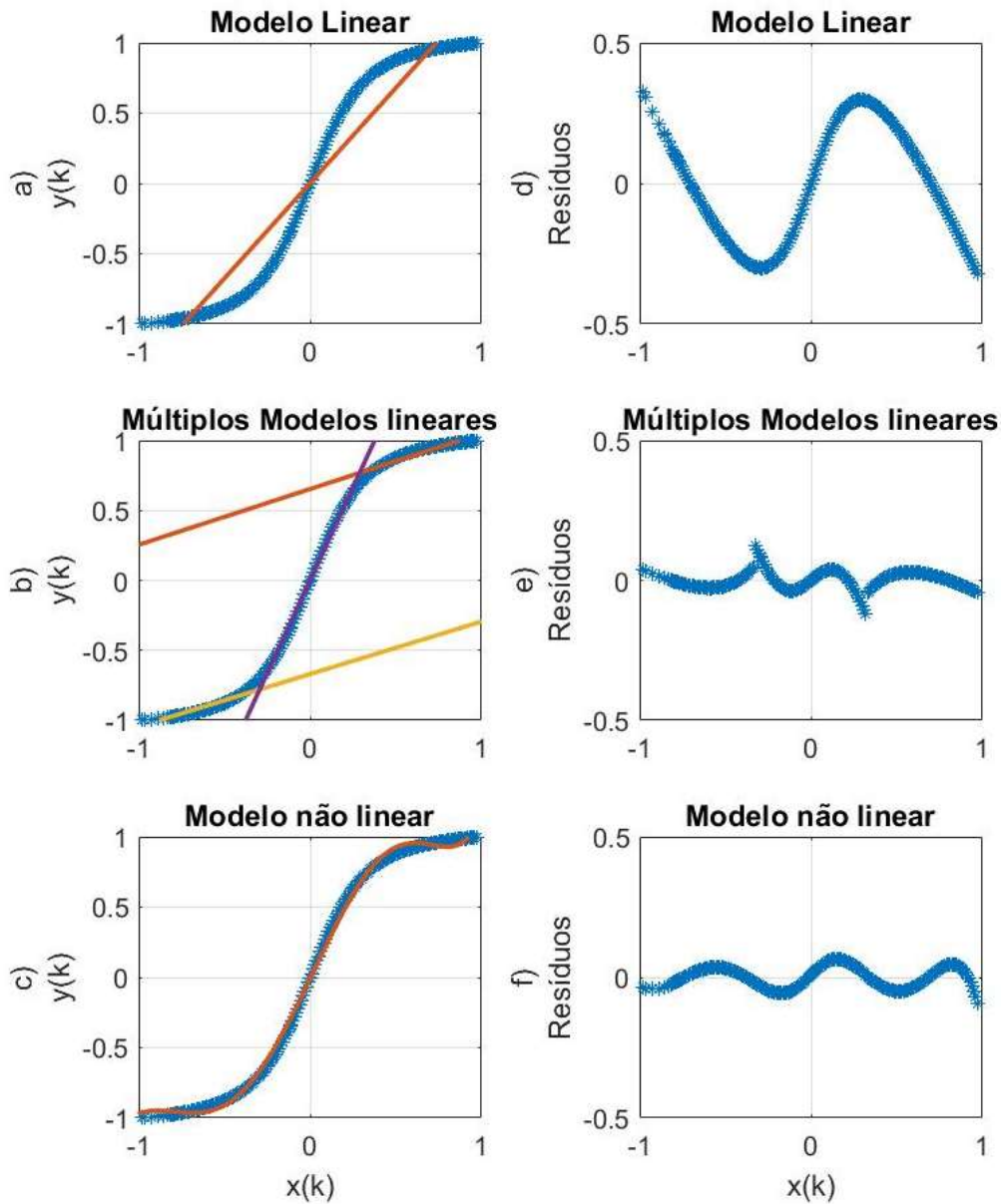


Figura 5 - Comparação de modelos lineares (a), múltiplos modelos lineares (b) e modelo polinomial. d) a f) mostram os resíduos dos modelos

Observa-se que o modelo linear da Figura 5a apresenta elevados valores absolutos dos resíduos (Figura 5d) quando comparado aos outros modelos, mesmo com o processo operando sem falhas. Ao utilizar-se deste modelo linear em uma abordagem de Análise Externa para detectar falhas, como descrito no item 2.2 e 2.3, estes elevados resíduos são considerados normais na etapa de treinamento, gerando baixa sensibilidade as falhas durante o monitoramento. Ou seja, os valores absolutos dos resíduos de treinamento devem apresentar média próxima a zero afim de melhorar a sensibilidade às falhas. Desta forma, espera-se que os métodos propostos neste trabalho, baseados em múltiplos

modelos lineares e modelos não lineares apresentem uma sensibilidade às falhas superior a metodologia de Análise Externa linear original.

Os métodos propostos, em conjunto com o método de Análise Externa linear proposto por Kano (2004), identificado por AEL, são aplicados a dois estudos de caso no capítulo 4, e são gerados os índices de falsos positivos (FP) e falsos negativos (FN) (Hossin, 2015) para a comparação entre as metodologias, sendo que FP é um índice que contabiliza os alarmes de falhas falsos e FN é um índice que contabiliza a quantidade de amostras que a metodologia considerou normal com o processo em falha, sendo que quanto menor seu valor, maior a sensibilidade às falhas.

Os níveis aceitáveis de FP estão associados ao nível de confiança usado para calcular o limiar estatístico da estatística  $T^2$ , que para este trabalho foi de 95%. Desta forma, valores próximos de 5% de FP ou alarmes falsos são aceitáveis.

A porcentagem de variância mantida pelas componentes principais em todos os métodos será de 90%.

Visando avaliar a generalização dos modelos gerados, é utilizado o método de validação cruzada *k-fold* (T. Hastie, 2001) em todos os métodos propostos com  $k = 5$ . Ou seja, as matrizes de dados de treinamento dos modelos têm suas amostras divididas em cinco conjuntos, sendo que quatro são usados para treinamento e um para validação. O procedimento de treino se repete por cinco vezes, alterando os conjuntos de treino e validação. São escolhidos os modelos associados aos conjuntos que minimizam a norma dos resíduos dos modelos de Análise Externa.

### **3.1 Análise Externa com Múltiplos Modelos Lineares - AEML**

O método proposto nesta seção faz uso de múltiplos modelos de Análise Externa linear e é identificado como AEML (Análise Externa com Múltiplos Modelos Lineares), onde cada modelo é utilizado em uma determinada região de operação. Os múltiplos modelos de Análise Externa são criados a partir do cálculo e classificação de submodelos de regressões lineares, a partir de uma janela móvel das amostras dos dados. A partir dos resíduos dos modelos de Análise Externa, o monitoramento de falhas é realizado utilizando PCA e estatística  $T^2$ .

O método proposto é dividido em duas partes: treinamento (Seção 3.1.1) e monitoramento de falhas (Seção 3.1.2).

### 3.1.1 Treinamento - AEML

Considere o conjunto de dados de treinamento  $\mathbf{X} = [\mathbf{G} \ \mathbf{H}] \in \mathbb{R}^{n \times (m_G + m_H)}$  de um determinado processo em funcionamento normal, composto por  $n$  amostras contendo as matrizes de variáveis externas  $\mathbf{G}$  e variáveis principais  $\mathbf{H}$ . Dividiremos a matriz  $\mathbf{X}$  em segmentos com  $q$  amostras em sequência. A partir destes segmentos, modelos de regressão com matrizes de coeficientes  $\mathbf{W}_k$  são calculados usando as matrizes  $\mathbf{G}_k$  e  $\mathbf{H}_k$ , com  $k = 1, \dots, \text{int}(n/q)$ , obtidas para cada segmento com  $q$  amostras, usando MQO:

$$\mathbf{W}_k = (\mathbf{G}_k^T \mathbf{G}_k)^{-1} \mathbf{G}_k^T \mathbf{H}_k \quad (18)$$

É essencial que a escolha do segmento contenha variações das variáveis externas (sinal rico), que são os sinais de entrada do modelo. Ou seja, o segmento deve possuir informações do processo em mais de um ponto de operação, para que seja possível capturar as relações entre as variáveis externas e principais.

Posteriormente as matrizes  $\mathbf{W}_k$  são convertidas para vetores linhas, conforme equações abaixo:

$$\text{de:} \quad \mathbf{W}_k = \begin{bmatrix} w_{11} & \cdots & w_{m_H 1} \\ \vdots & \ddots & \vdots \\ w_{m_G 1} & \cdots & w_{m_G m_H} \end{bmatrix} \quad (19)$$

$$\text{para:} \quad \mathbf{z}_k = [w_{11} \cdots w_{m_H 1} \cdots w_{m_G 1} \cdots w_{m_G m_H}] \quad (20)$$

O próximo passo é agrupar os vetores linhas em uma matriz  $\mathbf{Z}$  com  $n/q$  linhas (truncar se  $n/q$  não for inteiro):

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_k \\ \vdots \\ \mathbf{z}_{n/q} \end{bmatrix} \quad (21)$$

onde cada linha de  $\mathbf{Z}$  contém os coeficientes de regressão do modelo para um segmento de dados. O algoritmo de agrupamento *k-means* (D. Arthur, 2007) é então aplicado às

linhas da matriz  $\mathbf{Z}$  para agrupar modelos com coeficientes similares. O número de *clusters*, e consequentemente o número de modelos, é selecionado analisando a norma dos resíduos dos modelos de Análise Externa. Se o aumento do número de modelos não reduz a norma dos resíduos, não é mais necessário o aumento de *clusters*.

Definido o número de clusters, os dados associados a todos os modelos que foram agrupados são utilizados para recalculer o modelo de regressão do grupo, assegurando-se assim uma melhor representatividade dos dados do grupo pelo modelo.

Os resíduos de cada modelo de Análise Externa são normalizados e posteriormente utilizados para treinar um modelo PCA, conforme detalhado na seção 2.4. São armazenados a média e desvio padrão dos resíduos para que possam ser utilizados na normalização de novas amostras durante a fase de monitoramento de falhas.

Durante o monitoramento de falhas, quando uma nova amostra é disponibilizada, é necessário avaliar qual modelo de Análise Externa e PCA deverá ser usado. Para isto é utilizado um classificador do tipo *Random Forest* (Breiman, 2001) tendo como entrada as variáveis externas  $\mathbf{G}$  rotuladas na etapa de treinamento. É calculado a matriz de confusão do classificador usando 80% das amostras disponíveis para treinamento e 20% para validação, de forma a garantir que o classificador esteja conseguindo identificar corretamente os modelos a serem utilizados.

### 3.1.2 Monitoramento de falhas - AEML

Durante o monitoramento de falhas, a partir de novas amostras de variáveis externas, o classificador *Random Forest* identifica o modelo a ser usado.

Com o modelo de Análise Externa e PCA definido, os resíduos  $E$  são calculados utilizando a equação (5). Posteriormente os resíduos são normalizados pela média e desvio padrão calculados na etapa de treinamento. A partir da equação (9) a estatística  $T^2$  é calculada e comparada com o limiar estatístico calculado conforme equação (10) para detecção das falhas.

Um resumo dos passos do método será apresentado a seguir.

#### **Treinamento:**

1. Coleta de  $n$  amostras de  $m$  variáveis do processo em operação normal na matriz  $\mathbf{X}$
2. Divisão de  $\mathbf{X}$  em  $k$  conjuntos para realizar validação cruzada *k-fold*.  $k - 1$  conjuntos serão usados para geração dos modelos, que chamaremos de



$X_{treino}$ , e 1 conjunto será usado para validação dos modelos, que chamaremos de  $X_{val}$ .

3. Divisão de  $X_{treino}$  em segmentos com  $q$  amostras
4. Estimação dos parâmetros, via MQO, dos modelos dos segmentos, tendo as variáveis externas como entrada e principais como saída
5. Agrupamento dos modelos dos segmentos em  $N$  clusters usando  $k$ -means
6. Estimação dos parâmetros, via MQO, de um modelo para cada cluster usando os dados dos modelos agrupados
7. Cálculo dos resíduos usando dados de treinamento  $X_{treino}$  e modelos
8. Normalização dos resíduos
9. Geração de modelo PCA a partir dos resíduos normalizados
10. Cálculo do limiar estatístico
11. Treinamento de um classificador usando como entrada as variáveis externas rotuladas durante o agrupamento, contidas em  $X_{treino}$
12. Verificação da matriz de confusão do classificador
13. Rotulação das amostras de validação  $X_{val}$  utilizando o classificador
14. Cálculo dos resíduos usando dados de validação  $X_{val}$  e modelos selecionados pelo classificador
15. Cálculo da norma dos resíduos
16. Repetir os passos 3 a 15 por  $k$  vezes, usando todas as combinações possíveis de conjuntos de treinamento e validação
17. Escolher os modelos gerados a partir dos conjuntos de treinamento e validação com menor norma dos resíduos para serem usados na etapa de monitoramento

### **Monitoramento:**

1. Obtenção de nova amostra  $x$
2. Divisão das colunas de  $x$  em variáveis externas  $g$  e principais  $h$
3. Seleção do modelo usando o classificador a partir da amostra das variáveis externas
4. Cálculo do resíduo usando o modelo selecionado
5. Normalização do resíduo correspondente ao modelo
6. Teste estatístico do resíduo normalizado

### 3.2 Análise Externa com Múltiplos Modelos Lineares Baseado em RANSAC - AEML-R

Assim como o método AEML, esse método propõe o uso de múltiplos modelos de Análise Externa linear, onde cada modelo será utilizado em uma determinada região de operação. Os múltiplos modelos são criados a partir da metodologia consenso de amostra aleatória (*random sample consensus* – RANSAC). Esse método é identificado como AEML-R (Análise Externa com Múltiplos Modelos Lineares via RANSAC). Assim como no método AEML, a partir dos resíduos dos modelos, a detecção de falha é realizada utilizando PCA e estatística  $T^2$ .

O método proposto é dividido em duas partes: treinamento (Seção 3.2.1) e monitoramento de falhas (Seção 3.2.2).

#### 3.2.1 Treinamento - AEML-R

A metodologia RANSAC foi proposta por Fischler (1981) para modelar sistemas retirando a influência de amostras atípicas, que podem comprometer a qualidade do modelo. Para utilizar esta metodologia, algumas escolhas devem ser feitas:

1. Estrutura do modelo a ser gerado
2. Função de distância, associada ao erro de modelagem
3. Máxima distância permitida
4. Quantidade mínima de amostras para geração de um modelo
5. Número máximo de iterações
6. Número máximo de modelos

A partir das entradas definidas, esta metodologia busca iterativamente um modelo com a estrutura escolhida que maximize o número de amostras que apresentem distância menor que a máxima definida. Essas amostras são chamadas *inliers* e as amostras que possuem distância maiores que a máxima definida são chamadas de *outliers*. As distâncias são calculadas a partir da função de distância escolhida. Somente serão validados modelos que possuem um número mínimo de *inliers*.

A metodologia proposta nesta seção utiliza a metodologia RANSAC de forma iterativa para gerar os múltiplos modelos lineares. Inicialmente o algoritmo RANSAC é

utilizado para todas as amostras disponíveis. Posteriormente o algoritmo é utilizado novamente para os *outliers* do modelo anterior e assim sucessivamente.

A Figura 6 mostra este método aplicado ao exemplo de válvulas de controle detalhado no início do capítulo 3.

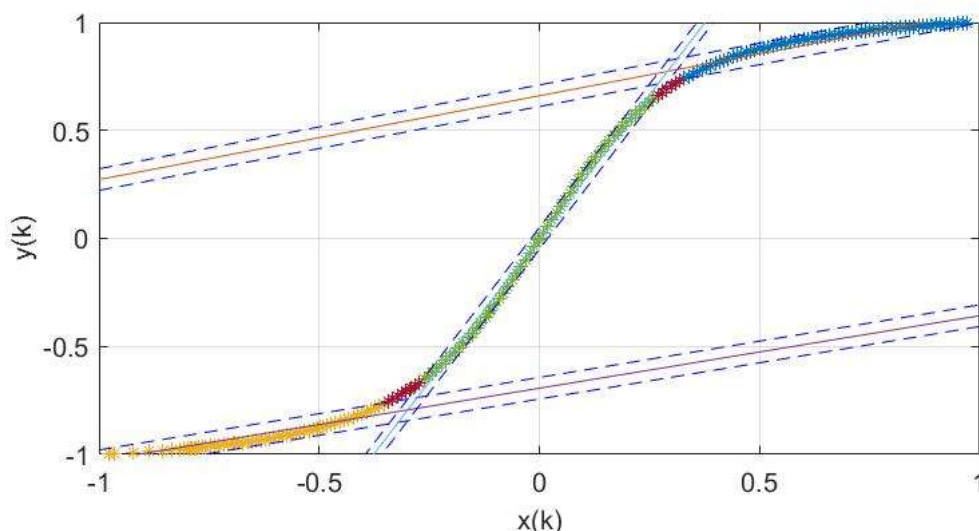


Figura 6 – Exemplo de múltiplos modelos lineares usando RANSAC para uma válvula de controle (posição x fluxo)

Na primeira iteração do algoritmo RANSAC, é gerado um modelo linear que possui as amostras em verde como *inliers*, identificado na Figura 6. O segundo modelo gerado possui as amostras em azul como *inliers* e o terceiro modelo possui as amostras em amarelo como *inliers*. As amostras em vermelho são *outliers* do último modelo e não respeitam a máxima distância permitida para nenhum modelo. As máximas distâncias permitidas são representadas pelas linhas tracejadas.

Neste trabalho utiliza-se uma estrutura de modelo baseada em MQO e a função de distância será a norma euclidiana do erro de modelagem, os mesmos parâmetros usados no exemplo da Figura 6.

A máxima distância permitida deve ser escolhida de acordo com a variabilidade das variáveis, que varia para cada aplicação, sendo que valores pequenos dificultam a geração dos modelos e valores elevados podem aumentar a norma dos resíduos dos modelos gerados e consequentemente diminuir a sensibilidade de detecção de falhas.

É essencial que a escolha da quantidade mínima de amostras para geração de um modelo seja feita de forma a garantir que o conjunto de dados usados contenha variações das variáveis externas (sinal rico), que são os sinais de entrada do modelo. Ou seja, o segmento deve possuir informações do processo em mais de um ponto de operação, para

que seja possível capturar as relações entre as variáveis externas e principais. Valores muito elevados dificultarão a geração de modelos.

O número máximo de iterações deve ser suficiente para que o algoritmo consiga encontrar os modelos de forma iterativa a partir dos parâmetros definidos.

O número máximo de modelos deve ser definido avaliando-se a quantidade de amostras que foram consideradas *outliers* do último modelo gerado, tendo em vista que estão associados as amostras que possui distância maior que a máxima permitida para todos os modelos.

Estas escolhas podem ser trabalhosas e em alguns casos a geração de gráficos como o da Figura 6, que ilustram a geração dos modelos, ajudam nas escolhas.

Os resíduos de cada modelo linear de Análise Externa são normalizados e posteriormente são utilizados para treinar um modelo PCA, conforme detalhado na seção 2.4. São armazenados a média e desvio padrão dos resíduos para serem utilizados na normalização de novas amostras durante a fase de monitoramento de falhas.

Assim como no método AEML, durante o monitoramento de falhas, quando uma nova amostra é disponibilizada, é necessário avaliar qual modelo de Análise Externa e PCA deverá ser usado. Para isto será utilizado um classificador do tipo *Random Forest* (Breiman, 2001) tendo como entrada as variáveis externas  $G$  rotuladas na etapa de treinamento. Será calculado a matriz de confusão do classificador usando 80% das amostras disponíveis para treinamento e 20% para validação, de forma a garantir que o classificador esteja conseguindo identificar corretamente os modelos a serem utilizados.

### 3.2.2 Monitoramento de falhas – AEML-R

Durante o monitoramento de falhas, a partir de novas amostras de variáveis externas, o classificador *Random Forest* identifica o modelo a ser usado.

Com o modelo de análise externa e PCA definido, os resíduos  $E$  são calculados utilizando a equação (5). Posteriormente os resíduos são normalizados pela média e desvio padrão calculados na etapa de treinamento. A partir da equação (9) a estatística  $T^2$  pode ser calculada e comparada com o limiar estatístico calculado conforme equação (10) para detecção das falhas.

Um resumo dos passos do método será apresentado a seguir.

#### **Treinamento:**

1. Coleta de  $n$  amostras de  $m$  variáveis do processo em operação normal na matriz  $\mathbf{X}$
2. Divisão de  $\mathbf{X}$  em  $k$  conjuntos para realizar validação cruzada  $k$ -fold.  $k - 1$  conjuntos serão usados para geração dos modelos, que chamaremos de  $\mathbf{X}_{treino}$ , e 1 conjunto será usado para validação dos modelos, que chamaremos de  $\mathbf{X}_{val}$ .
3. Aplicação de RANSAC a  $\mathbf{X}_{treino}$ , gerando um modelo de regressão linear via MQO
4. Aplicação de RANSAC aos *outliers* do modelo gerado anteriormente até que não seja mais possível a geração de um novo modelo
5. Cálculo dos resíduos usando dados de treinamento  $\mathbf{X}_{treino}$  e modelos
6. Normalização dos resíduos
7. Cálculo do modelo PCA a partir dos resíduos normalizados
8. Cálculo do limiar estatístico
9. Treinamento de um classificador usando como entrada as variáveis externas rotuladas durante o agrupamento, contidas em  $\mathbf{X}_{treino}$
10. Verificação da matriz de confusão do classificador
11. Rotulação das amostras de validação  $\mathbf{X}_{val}$  utilizando o classificador
12. Cálculo dos resíduos usando dados de validação  $\mathbf{X}_{val}$  e modelos selecionados pelo classificador
13. Cálculo da norma dos resíduos
14. Repetir os passos 3 a 13 por  $k$  vezes, usando todas as combinações possíveis de conjuntos de treinamento e validação
15. Escolher os modelos gerados a partir dos conjuntos de treinamento e validação com menor norma dos resíduos para serem usados na etapa de monitoramento

### **Monitoramento:**

1. Obtenção de nova amostra  $\mathbf{x}$
2. Divisão das colunas de  $\mathbf{x}$  em variáveis externas  $\mathbf{g}$  e principais  $\mathbf{h}$
3. Seleção do modelo usando o classificador a partir da amostra das variáveis externas
4. Cálculo do resíduo usando o modelo selecionado
5. Normalização do resíduo correspondente ao modelo
6. Teste estatístico do resíduo normalizado

### 3.3 Análise Externa não linear

Nesta seção são propostos três métodos de Análise Externa não linear utilizando somente um modelo baseado em regressão polinomial, identificado como AEPOL, rede neural de regressão geral (*General Regression Neural Network* - GRNN), identificado como AEGRNN e *Random Forest*, identificado como AERF. A partir dos resíduos do modelo de Análise Externa, a detecção de falha é realizada utilizando PCA e estatística  $T^2$ .

Os métodos propostos são divididos em duas partes: treinamento (Seção 3.3.1) e monitoramento de falhas, (Seção 3.3.2).

#### 3.3.1 Treinamento – Análise Externa não linear

O método AEPOL gera o modelo de Análise Externa usando o método de MQO, conforme equação (3), porém, é realizada uma modificação na matriz  $\mathbf{G}$ , incluindo novas colunas como sendo potências das variáveis, conforme a ordem definida para o polinômio. Com esta modificação da matriz de variáveis externas, obtém-se um modelo de regressão polinomial. A decisão sobre aumentar a ordem da regressão polinomial é feita analisando a norma euclidiana dos resíduos. Se o aumento da ordem do polinômio não reduz a norma dos resíduos, não é mais necessário o aumento de ordem.

O método AEGRNN gera o modelo de Análise Externa usando GRNN, proposto por Specht (1991), a partir da matriz de variáveis externas  $\mathbf{G}$  como entradas e variáveis principais  $\mathbf{H}$  como saídas. Este modelo possui somente um parâmetro que está ligado ao erro de modelagem permitido. Valores muito baixos desse parâmetro podem gerar *overfit*, que é a tentativa de modelagem dos ruídos inerentes a aplicação e valores elevados podem aumentar a norma dos resíduos. Neste trabalho o parâmetro é escolhido de forma a minimizar a norma dos resíduos, mantendo os falsos positivos dos dados de validação condizentes com o nível de confiança escolhido para calcular o limiar estatístico, de forma a evitar *overfit*.

O método AETB gera o modelo de Análise Externa usando *Random Forest*, proposto por Breiman (2001), a partir da matriz de variáveis externas  $\mathbf{G}$  como entradas e variáveis principais  $\mathbf{H}$  como saídas. São ajustados como parâmetros deste modelo a quantidade de árvores e a profundidade dos ramos de cada árvore. O aumento da profundidade dos ramos de cada árvore pode gerar *overfit*, assim como o número reduzido de quantidade de árvores. Neste trabalho estes parâmetros são ajustados de forma a minimizar a norma

dos resíduos, mantendo os falsos positivos dos dados de validação condizentes com o nível de confiança escolhido para calcular o limiar estatístico, de forma a evitar o *overfit*.

Os resíduos de cada modelo não linear de Análise Externa são normalizados e posteriormente são utilizados para treinar um modelo PCA, conforme detalhado na seção 2.4. São armazenados as médias e desvios padrões dos resíduos para que possam ser utilizados na normalização de novas amostras durante a fase de monitoramento de falhas.

### **3.3.2 Monitoramento de falhas – Análise Externa não linear**

Quando novas amostras são disponibilizadas, durante o monitoramento de falhas, os resíduos  $E$  da Análise Externa são calculados utilizando a equação (5). Posteriormente os resíduos são normalizados pela média e desvio padrão calculados na etapa de treinamento. A partir da equação (9) a estatística  $T^2$  é calculada e comparada com o limiar estatístico calculado conforme equação (10).

## 4 Estudos de caso

Neste capítulo, as metodologias propostas no capítulo 3 em conjunto com a metodologia de Análise Externa proposta por Kano (2004) são aplicadas a um simulador de reator contínuo do tipo tanque agitado (*Continuous Stirred Tank Reactor - CSTR*) (OYELEYE, 1989), (FINCH, 1989) e em um monitoramento real de vibrações de um ventilador de processo de grande porte, usado em um forno de pelotização de minério de ferro da mineradora Vale S.A.

### 4.1 Reator contínuo do tipo tanque agitado (*Continuous Stirred Tank Reactor - CSTR*)

O fluxograma do processo CSTR usado neste estudo de caso é mostrado na Figura 7.

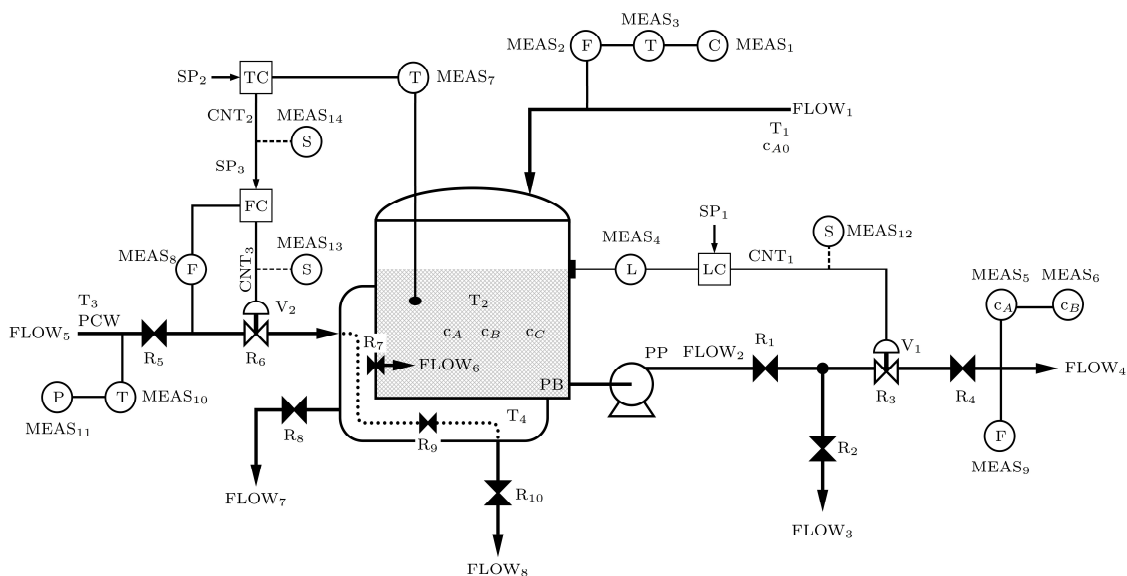


Figura 7 - Fluxograma do processo CSTR

O reagente A com concentração  $c_{A0}$  a temperatura  $T_1$  alimenta o reator com vazão  $FLOW_1$ , onde duas reações em paralelo ocorrem  $A \rightarrow B$  e  $A \rightarrow C$ . A primeira reação é dominante e é exotérmica e a segunda endotérmica, sendo que o balanço energético geral é exotérmico, gerando o aumento da temperatura do reator (MEAS7). Os produtos B e C e o restante do reagente A saem do tanque e são bombeados com vazão  $FLOW_4$  (quando  $FLOW_3$  é igual a zero) e concentração  $c_A$  e  $c_B$  (concentração de  $c_C$  é ignorada). O controle de nível do reator é feito através da definição de  $SP_1$ , que modula a válvula de controle



de produto final V1. O controle de temperatura do reator é feito através da definição do SP<sub>2</sub>, que controla uma malha em cascata que modula a válvula de controle de água de refrigeração da jaqueta do reator V2. A temperatura, pressão e vazão de entrada água de refrigeração são as variáveis MEAS<sub>10</sub>, MEAS<sub>11</sub> e FLOW<sub>5</sub>, respectivamente. 14 variáveis de processos e 2 *Set points* compõe este sistema, conforme descrito na Tabela 1.

Tabela 1 - Variáveis de processo CSTR

Variável	Descrição	Observação
MEAS <sub>1</sub>	Concentração alimentação	-
FLOW <sub>1</sub>	Vazão alimentação	-
MEAS <sub>3</sub>	Temperatura alimentação	-
MEAS <sub>4</sub>	Nível reator	Variável principal
MEAS <sub>5</sub>	Concentração produto A	Variável principal
MEAS <sub>6</sub>	Concentração produto B	Variável principal
MEAS <sub>7</sub>	Temperatura Reator	Variável principal
FLOW <sub>5</sub>	Vazão água refrigeração	Variável principal
FLOW <sub>4</sub>	Vazão produto	Variável principal
MEAS <sub>10</sub>	Temp. entrada água refrigeração	-
MEAS <sub>11</sub>	Pressão entrada água refrigeração	Variável externa
MEAS <sub>12</sub>	Saída controlador nível	Variável principal
MEAS <sub>13</sub>	Saída controlador vazão	Variável principal
MEAS <sub>14</sub>	<i>Set point</i> controlador de vazão	Variável principal
SP <sub>1</sub>	<i>Set point</i> nível reator	-
SP <sub>2</sub>	<i>Set point</i> temperatura reator	-

Neste estudo de caso, utiliza-se a pressão de água de refrigeração (MEAS<sub>11</sub>) como variável externa, sendo que mudanças nesta variável modificam o ponto de operação do processo. Esta variável é escolhida por ter relações não lineares com outras variáveis, de forma a representar de maneira mais adequada processos industriais reais. As variáveis principais são identificadas na Tabela 1, sendo que variáveis de entrada do processo que são independentes da variável externa não são utilizadas.

O conjunto de dados de treinamento e avaliação de falsos positivos são gerados a partir de uma simulação sem a presença de falhas, com 22 variações em MEAS<sub>11</sub>, totalizando 7508 amostras com tempo de amostragem de 1 minuto. A Figura 8 mostra os dados de treinamento da variável externa MEAS<sub>11</sub>.

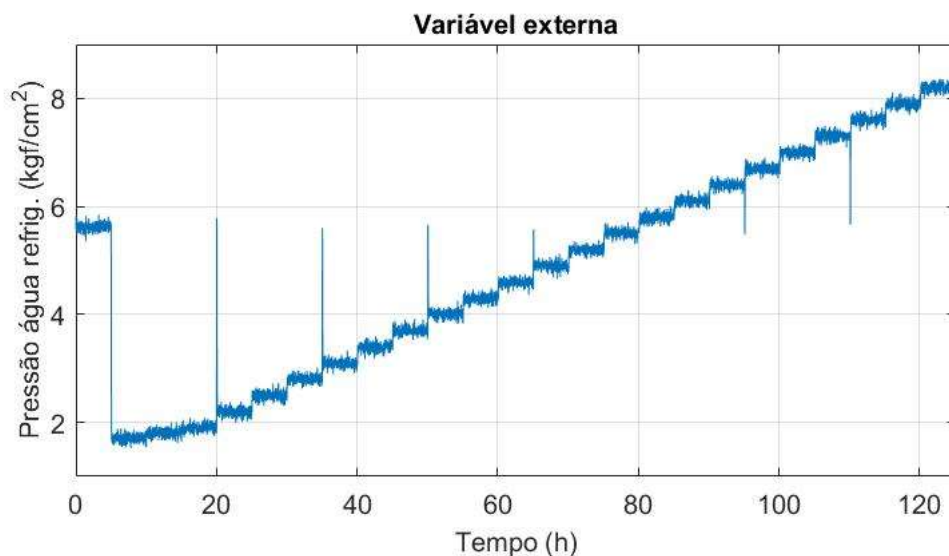


Figura 8 – Variações da variável externa MEAS<sub>11</sub> para treinamento

A seguir são apresentados os parâmetros utilizados em cada uma das metodologias a serem comparadas:

- AEML
  - Número de modelos: 4
  - Janela de dados q: 300 amostras
- AEML-R
  - Estrutura do modelo: MQO
  - Função de distância: norma euclidiana dos resíduos
  - Distância máxima permitida: 3,8
  - Quantidade mínima de amostras para geração de um modelo: 3% das amostras do conjunto de dados de treinamento
  - Número máximo de iterações 1.000
  - Número máximo de modelos: 4
- AEPOL
  - Ordem do polinômio: 5
- AEGRNN
  - Parâmetro de espalhamento: 200
- AERF
  - Quantidade máxima de árvores de decisão: 10
  - Profundidade das árvores: 25

Como descrito no capítulo 3, é utilizada o método de validação cruzada *k-fold* com *k* igual a 5 na etapa de treinamento, de forma a escolher os modelos que minimizam a norma dos resíduos dos dados de validação. Ou seja, 7508 amostras de treinamento são aleatoriamente divididas em *k* conjuntos, sendo que *k* – 1 conjuntos são utilizados para geração dos modelos e 1 conjunto é usado para validação. Repete-se este procedimento até obter todas as combinações possíveis, alternando os conjuntos de validação e treinamento. A Figura 9 mostra a norma total dos resíduos dos modelos de Análise Externa de todos os métodos, gerados a partir da validação cruzada.

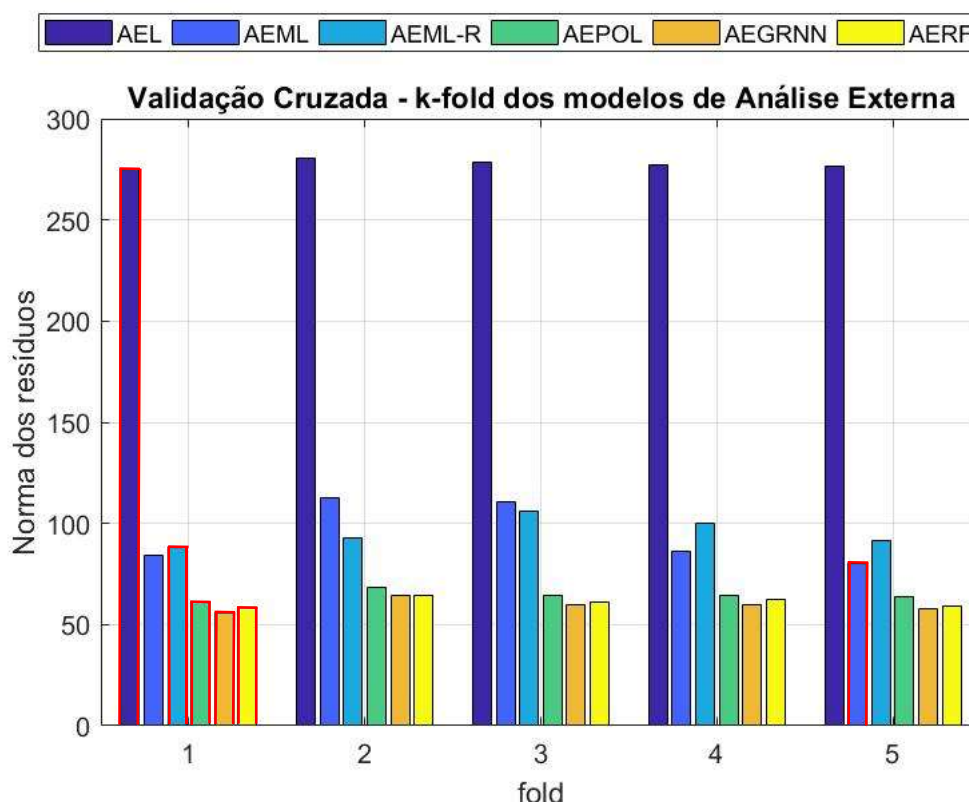


Figura 9 – Validação cruzada: Norma dos resíduos de Análise Externa aplicados aos conjuntos de dados de validação

A norma total dos resíduos da etapa de treinamento, mostrada na Figura 9, está diretamente associada à sensibilidade as falhas dos métodos. Isto ocorre devido as falhas serem detectadas a partir do aumento dos resíduos dos modelos de Análise Externa. Se, durante o treinamento, resíduos elevados são considerados normais, somente desvios ainda maiores serão detectados como falha. A norma elevada dos resíduos da etapa de treinamento gera média e/ou desvio padrão elevados, que por sua vez são utilizados para

normalização de novas amostras na etapa de monitoramento, que acabam minimizando o efeito de falhas nos resíduos normalizados.

Conforme a Figura 9, verifica-se que a metodologia tradicional de Análise Externa (AEL) gera resíduos relativamente elevados quando comparados com as outras metodologias propostas neste trabalho.

Verifica-se que as metodologias que utilizam múltiplos modelos lineares (AEML e AEML-R) são mais sensíveis a escolha dos dados de treinamento, pois há maiores variações das normas dos resíduos quando se utilizam diferentes conjuntos de dados de treino e validação. Desta forma, é essencial o uso de validação cruzada para se obter bons modelos para estes métodos.

Para este estudo de caso são escolhidos os modelos associados aos conjuntos de dados de validação cruzada que obtiverem as menores normas dos resíduos para cada uma das metodologias, sendo representados por um contorno vermelho nas barras da Figura 9.

Para validar se o classificador *Random Forest* está selecionando corretamente os modelos dos métodos AEML e AEML-R, calcula-se a matriz de confusão, mostradas nas Figuras 10 e 11.

**Matriz de confusão classificador AEML**

Resposta do classificador	Modelo 1	510 42.5%	13 1.1%	1 0.1%	9 0.8%	95.7% 4.3%
	Modelo 2	24 2.0%	151 12.6%	9 0.8%	1 0.1%	81.6% 18.4%
	Modelo 3	0 0.0%	9 0.8%	219 18.3%	10 0.8%	92.0% 8.0%
	Modelo 4	13 1.1%	0 0.0%	11 0.9%	220 18.3%	90.2% 9.8%
		93.2% 6.8%	87.3% 12.7%	91.3% 8.8%	91.7% 8.3%	91.7% 8.3%
		Modelo 1	Modelo 2	Modelo 3	Modelo 4	
Dados rotulados						

Figura 10 – CSTR: Matriz de confusão classificador *Random Forest* AEML

Na figura 10, as células diagonais mostram o número e a porcentagem de amostras que foram classificadas corretamente, ou seja, para estas amostras o modelo correto foi selecionado pelo classificador. Por exemplo, a primeira célula da diagonal mostra que 510 amostras associadas ao modelo 1 são corretamente classificadas. As 510 amostras correspondem a 42,5% de todas as amostras disponíveis para validação do classificador.

A célula da segunda linha e primeira coluna mostra que 24 amostras pertencentes ao modelo 1 foram incorretamente classificadas como pertencentes ao modelo 2 e isso corresponde a 2% de todas as amostras disponíveis para classificação.

A somatória das amostras da primeira linha identifica todas as amostras classificadas como pertencentes ao modelo 1, 95,7% foram classificadas corretamente e 4,3% não. A somatória das amostras da primeira coluna identifica todas as amostras pertencentes ao modelo 1, 93,2% destas amostras foram classificadas corretamente e 6,8% não.

No geral, 91,7% das previsões estão corretas e 8,3% estão erradas (célula com fundo azul), indicando que o classificador do método AEML está funcionando de forma satisfatória.

**Matriz de confusão classificador AEML-R**

Resposta do classificador	Modelo 1	912 76.9%	1 0.1%	0 0.0%	1 0.1%	99.8% 0.2%
	Modelo 2	2 0.2%	164 13.8%	22 1.9%	9 0.8%	83.2% 16.8%
	Modelo 3	0 0.0%	29 2.4%	29 2.4%	3 0.3%	47.5% 52.5%
	Modelo 4	3 0.3%	6 0.5%	3 0.3%	2 0.2%	14.3% 85.7%
		99.5% 0.5%	82.0% 18.0%	53.7% 46.3%	13.3% 86.7%	93.3% 6.7%
		Modelo 1	Modelo 2	Modelo 3	Modelo 4	

Dados rotulados

Figura 11 – CSTR: Matriz de confusão classificador *Random Forest* AEML-R

Observa-se na Figura 11, que 93,3% das previsões estão corretas e 6,7% estão erradas, indicando que o classificador do método AEML-R está funcionando de forma satisfatória.

A Figura 12 mostra a estatística  $T^2$  de Hotelling e os FP (Falsos Positivos) de todas as metodologias aplicados aos dados de validação.

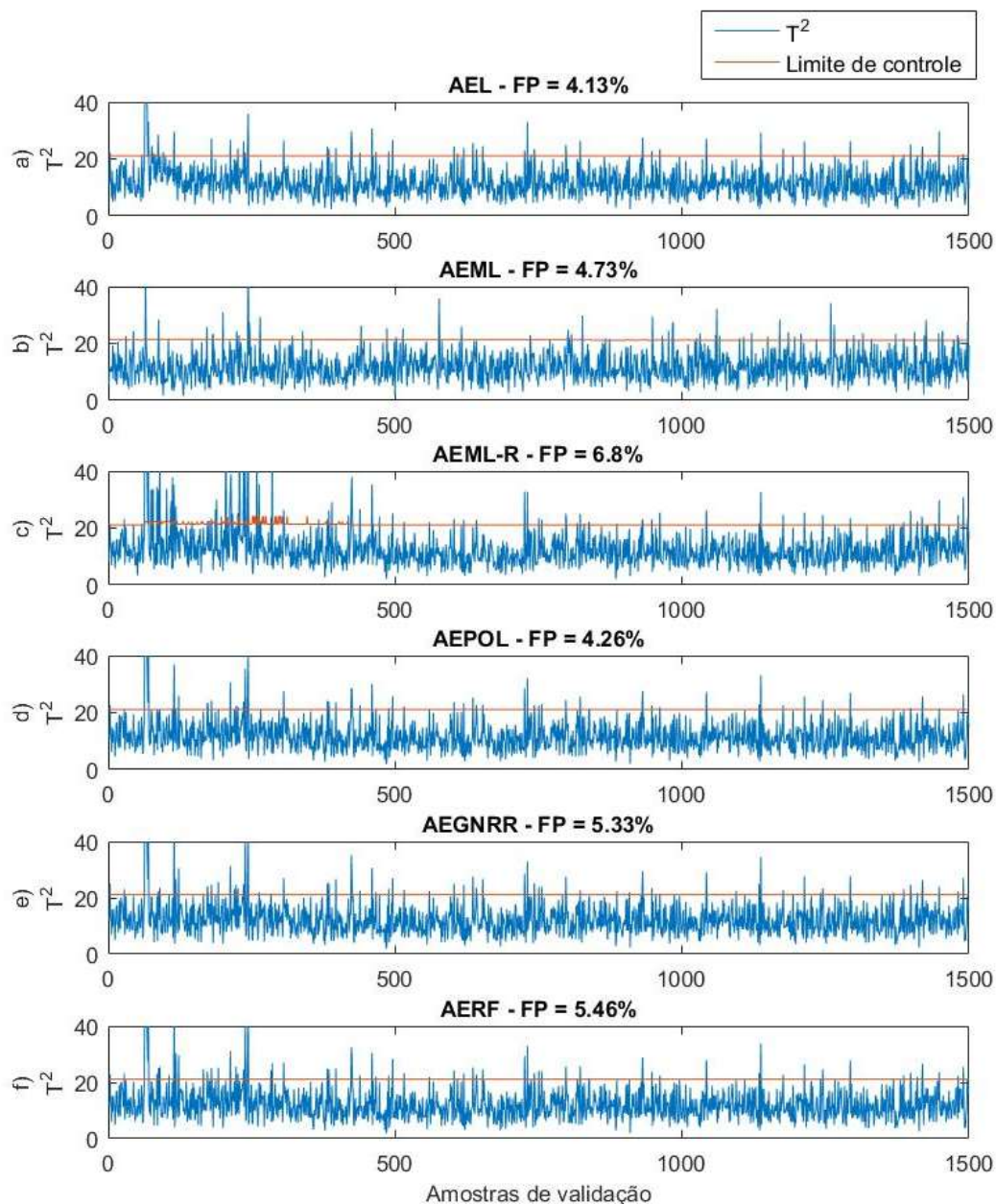


Figura 12 – Estatística  $T^2$  e avaliação de falsos positivos aplicados aos dados de validação. O gráfico a) mostra os resultados da metodologia AEL, b) AEML, c) AEML-R, d) AEPOL, e) AEGRNN e f) AERF

Como pode ser observado na Figura 12, para todas as metodologias os FP calculados são próximos ao valor esperado de 5%, tendo em vista que o limite de controle estatístico, em vermelho, foi calculado com o nível de confiança de 95%. As grandes diferenças entre as metodologias irão se manifestar no índice de falsos negativos, associado a sensibilidade a detecção de falhas.

O simulador do processo CSTR utilizado foi programado para simular 22 falhas de processo distintas, conforme descrito na Tabela 2.

Tabela 2 – Falhas do processo CSTR

Falha	Descrição
1	Sem falha
2	Bloqueio da saída do tanque
3	Bloqueio da jaqueta de refrigeração
4	Vazamento da jaqueta para o ambiente
5	Vazamento da jaqueta para o tanque
6	Vazamento na bomba
7	Perda de pressão da bomba
8	Baixa troca de calor na superfície da jaqueta
9	Fonte de aquecimento externo
10	Energia de ativação da reação primária
11	Energia de ativação da reação secundária
12	Vazão de alimentação anormal
13	Temperatura de alimentação anormal
14	Concentração de alimentação anormal
15	Temperatura de água de refrigeração anormal
16	Pressão de água de refrigeração anormal
17	Pressão de saída da jaqueta anormal
18	Pressão anormal de saída do reator
19	<i>Set point</i> anormal de nível
20	<i>Set point</i> anormal de temperatura
21	Válvula 1 presa
22	Válvula 2 presa
23	Falha de sensores (variáveis de processo)

Para avaliação dos FN (Falsos Negativos) utiliza-se a falha 3, que simula um aumento de restrição ao fluxo de água de refrigeração da jaqueta, reduzindo a troca de calor com o reator. Essa falha é escolhida, pois está diretamente relacionada à pressão de água de refrigeração ( $MEAS_{11}$ ), que é a variável externa neste estudo de caso.

Para comparar a capacidade de detecção de falhas das metodologias ao longo de toda a faixa de operação, a mesma falha é aplicada em 15 simulações com pontos de operação distintos, totalizando 30015 amostras, com tempo de amostragem de 1 minuto. Todas as falhas são iniciadas na amostra 200 de cada simulação. Para permitir a visualização de todos os resultados, os dados das 15 simulações são concatenados.



A Figura 13 mostra a estatística  $T^2$  e a avaliação de FN aplicados aos dados concatenados das simulações com falhas, assim como o gráfico da variável externa MEAS<sub>11</sub>, usada para modificar os pontos de operação do processo.

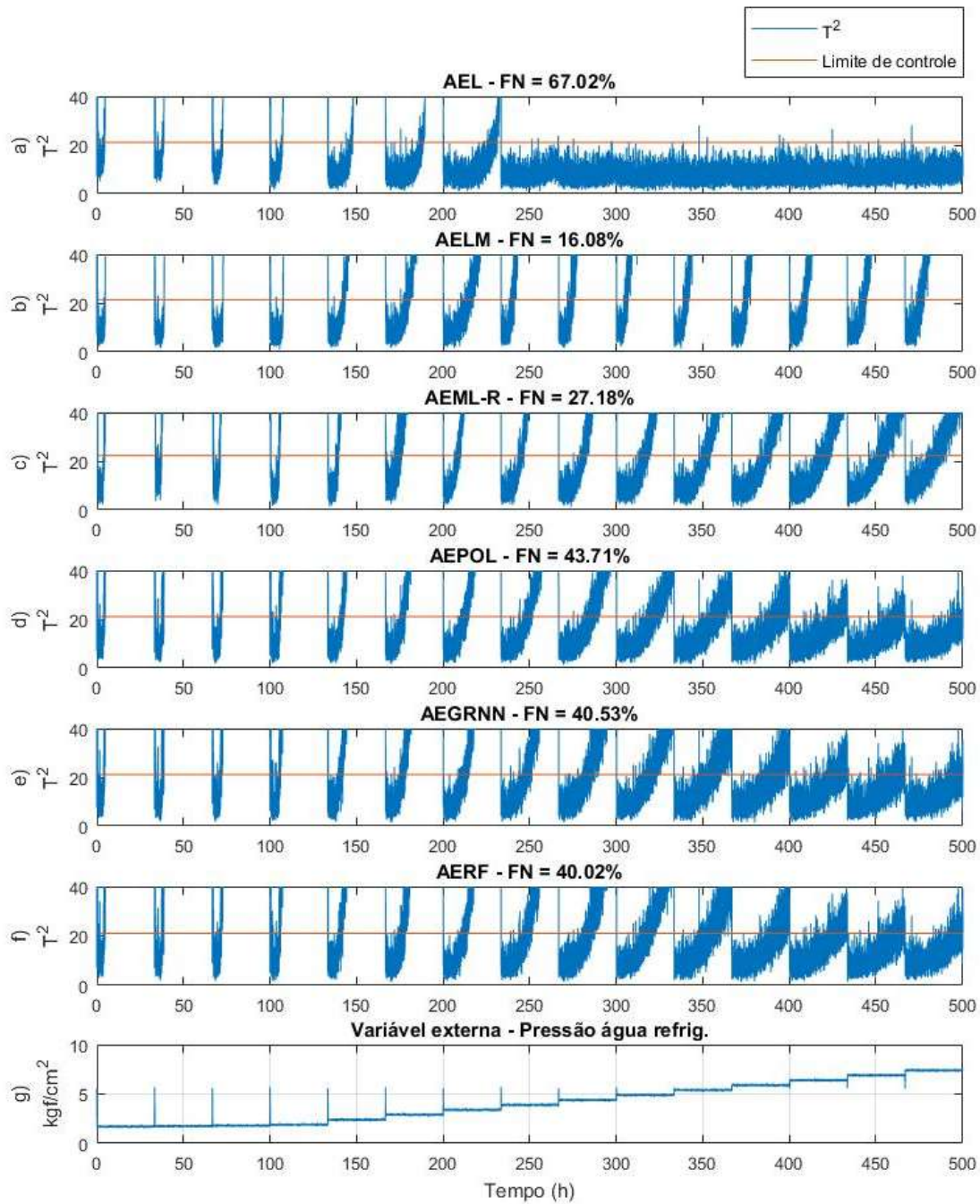


Figura 13 - Estatística  $T^2$  e avaliação de falsos negativos aplicados aos dados de falha em diferentes pontos de operação. O gráfico a) mostra os resultados da metodologia AEL, b) AELM, c) AELM-R, d) AEPOL, e) AEGRNN, f) AERF e g) mostra as variações da variável externa MEAS<sub>11</sub>



Analisando a Figura 13a, verifica-se que a metodologia AEL somente consegue detectar falhas nos primeiros 7 pontos de operação, pois a partir do ponto de operação 8 a estatística  $T^2$  não mais ultrapassa, de maneira estável, o limite de controle estatístico. Isso é um reflexo dos elevados resíduos do modelo de Análise Externa na etapa de treinamento. Para esta aplicação, que possui não linearidades, esta metodologia se mostrou insensível às falhas, não sendo aconselhado o seu uso.

Todas as outras metodologias se mostraram capazes de detectar as falhas em todos os pontos de operação. Conforme as Figuras 13b-f, o índice de falsos negativos mostrou que a metodologia AEML possui melhor resultado (16,08%), seguida das metodologias AEML-R (27,18%), AERF (40,02%), AEGRNN (40,53%) e AEPOL (43,71%). Valores baixos de falsos negativos indicam um menor tempo para detecção de falhas, o que pode gerar uma ação corretiva mais ágil, diminuindo os danos ao processo.

A Figura 14 mostra dos valores estratificados de FN por ponto de operação.

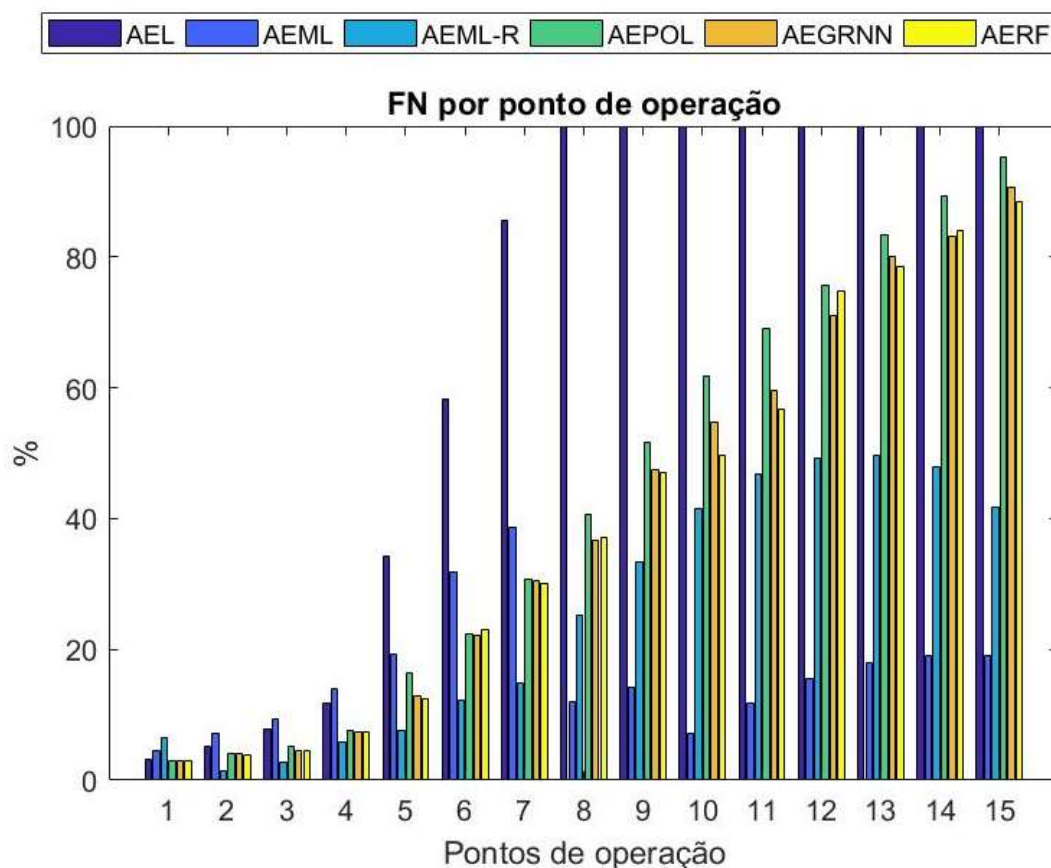


Figura 14 – FN estratificados por ponto de operação

Avaliando a Figura 14, verifica-se que em cada ponto de operação uma metodologia se sobressaiu, porém, a metodologia AEML foi a única que apresentou FN menor que 20% a partir do ponto de operação 8.

## **4.2 Monitoramento de vibrações de um ventilador de forno de pelotização**

O processo de pelotização de minério de ferro é um processo de aglomeração de finos de minério para geração de pelotas e exige um tratamento térmico, visando o aumento da resistência mecânica do produto para atender o processo posterior de geração de aço em siderúrgicas. O forno de pelotização é do tipo grelha móvel que recebe uma camada de pelotas e se movimenta ao longo do forno, passando por grupos de vários queimadores de combustíveis. A queima das pelotas é realizada através de fluxos de gases quentes que passam entre as pelotas, que são gerados por grandes ventiladores de processo, que também são utilizados para obter um reaproveitamento de gases quentes dentro do forno, de forma a diminuir o custo energético, assim como fornecer o oxigênio necessário para a queima do combustível.

Os ventiladores de processo são grandes máquinas, que possuem acionamento por inversores de frequência e motores elétricos de alta tensão, na faixa de 2-8MW de potência, e possuem muitos componentes mecânicos de alto valor. O funcionamento do forno e consequentemente da planta como um todo depende do funcionamento dos ventiladores de processo. Desta forma, são equipamentos críticos dentro do processo de pelotização, sendo que uma falha em um componente mecânico, além do custo do próprio componente, geram perdas econômicas elevadas quando consideramos o tempo para troca do componente, devido à parada de produção.

Os ventiladores mais comuns e que são objetos de estudo nesta seção, são do tipo radial, conforme mostrado na Figura 15. Possuem 4 mancais, um para o lado não acoplado (LNA) do ventilador, um para o lado acoplado (LA) do ventilador, uma para o mancal LA do motor e outro para o mancal LNA do motor. Estes mancais são lubrificados a óleo, que possui função de lubrificação e refrigeração dos mancais.

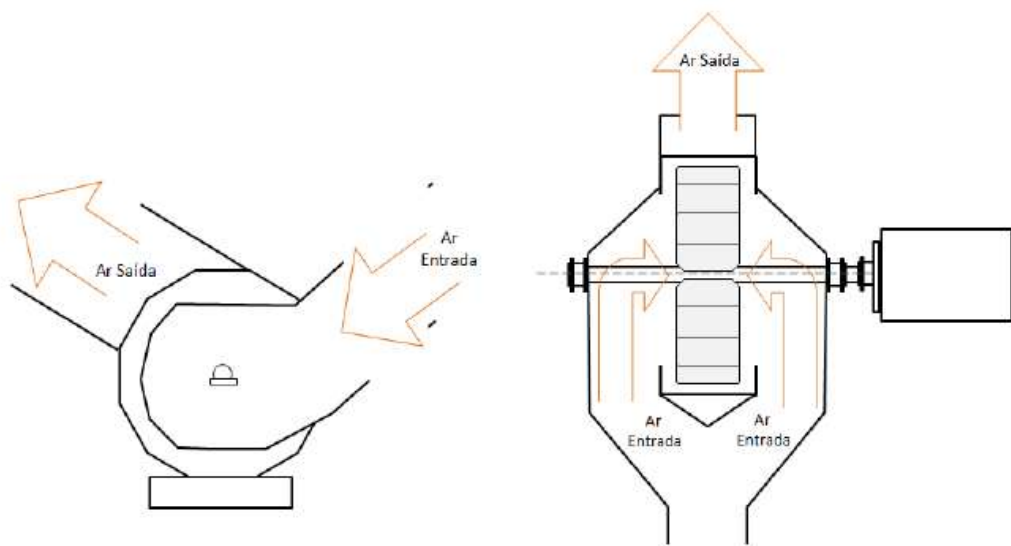


Figura 15 – Funcionamento de um ventilador de processo radial

Devido à criticidade destes equipamentos, eles possuem vários instrumentos de proteção, sendo os medidores de vibração dos mancais os mais importantes, visto que as falhas mais críticas afetam estas variáveis. Estes medidores medem a velocidade de vibração através do valor eficaz (*root mean square* - RMS), com faixa de medição de 0 a 25 mm/s. Este medidor é configurado com filtro na faixa de frequência de 10 a 1000Hz, para ser possível avaliar a severidade de vibração conforme NBR 10082.

Os sinais gerados pelos medidores de temperatura de mancais foram inicialmente utilizados, porém, distúrbios não medidos que afetavam estas variáveis geraram o aumento de alarmes falsos. Desta forma, somente os medidores de vibração são utilizados.

Neste estudo de caso são aplicadas as metodologias propostas na seção 3, em conjunto com o método AEL, no monitoramento de vibrações de mancais do ventilador 5Q7VT de uma usina de pelletização localizada em Vitória-ES, da mineradora Vale S. A..

A Figura 16 mostra a tela de detalhes deste ventilador, no sistema de supervisão da planta, identificando todas as variáveis de proteção deste equipamento, incluindo as variáveis que serão monitoradas neste estudo de caso.

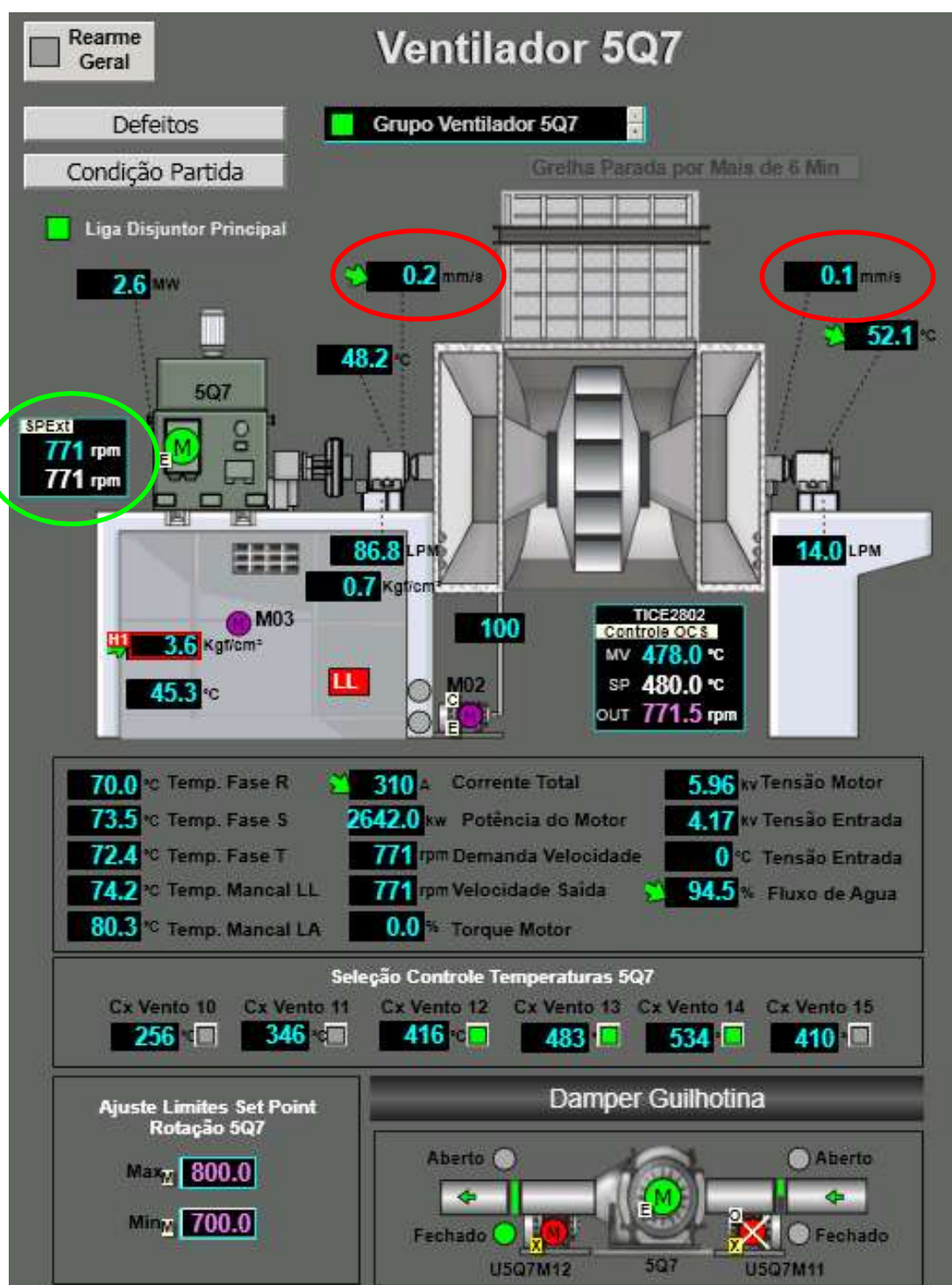


Figura 16 - Detalhes do sistema supervisor do ventilador 5Q7VT

Este ventilador possui inversor de frequência para modificar a rotação, que é o elemento final de controle de uma malha de controle de temperatura dentro do forno. Os pontos de operação das variáveis de proteção dos ventiladores são definidos pela rotação, indicada com um círculo verde na Figura 16, e por isso ela é definida como variável

externa. As vibrações do mancal acoplado e não acoplado do ventilador, identificadas em vermelho na Figura 16, são as variáveis principais a serem monitoradas.

A seleção de dados é realizada a partir do histórico de falhas de manutenção da empresa Vale S.A.. Em 17/05/2014 houve uma falha crítica no ventilador 5Q7VT, devido a avaria na ponta do eixo do rotor e no casquilho, que é um componente que é montado entre o mancal e o rotor do equipamento. O tempo para reparo gerou 210,23 horas de parada de produção de uma usina. São selecionados 2 dias de dados anteriores a falha, com tempo de amostragem de 10 segundos para treinamento das metodologias e avaliação dos índices FP e FN.

O objetivo do monitoramento é detectar falhas, através da análise do comportamento dos sinais de vibração, de forma a identificar anomalias em componentes mecânicos antes que a falha se agrave, diminuindo ou até mesmo evitando perdas financeiras para a empresa.

A Figura 17 mostra os dados da variável externa (rotação) e variáveis principais (vibrações) utilizados para treinamento e avaliação de FP.

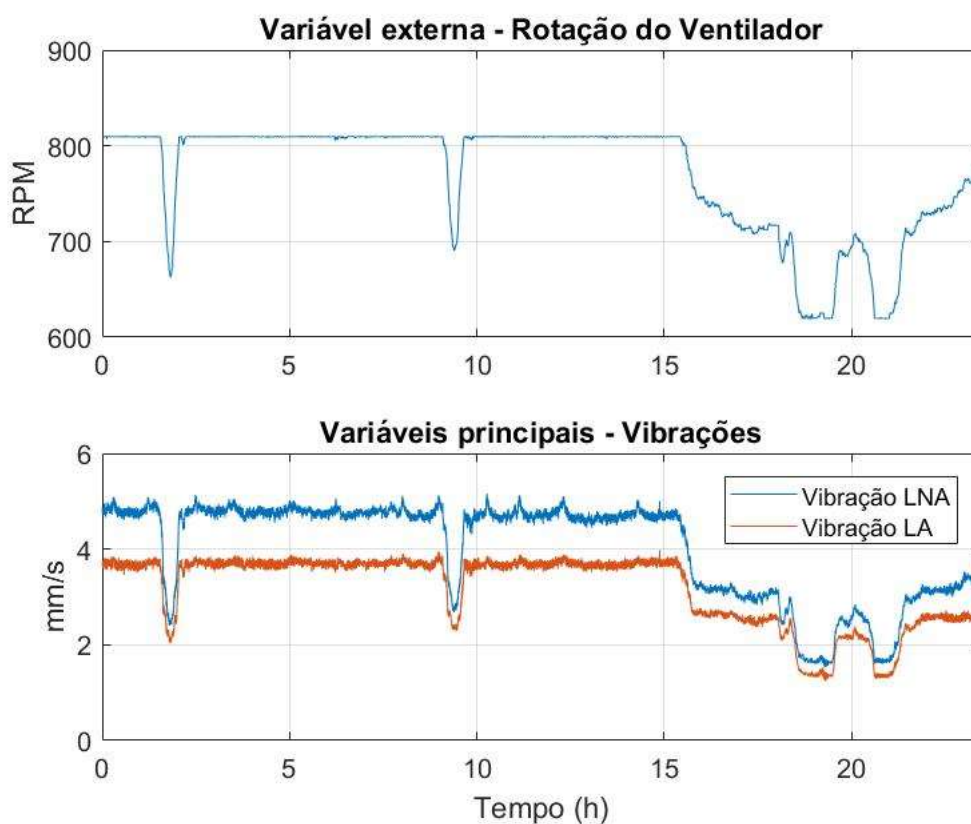


Figura 17 – Dados de rotação e vibração usados na etapa de treinamento e avaliação de FP

A Figura 18 mostra os gráficos de dispersão entre a rotação e vibrações em conjunto com uma reta gerada a partir da regressão linear via MQO. Observa-se que as relações entre as variáveis possuem não linearidades expressivas para a vibração LNA.

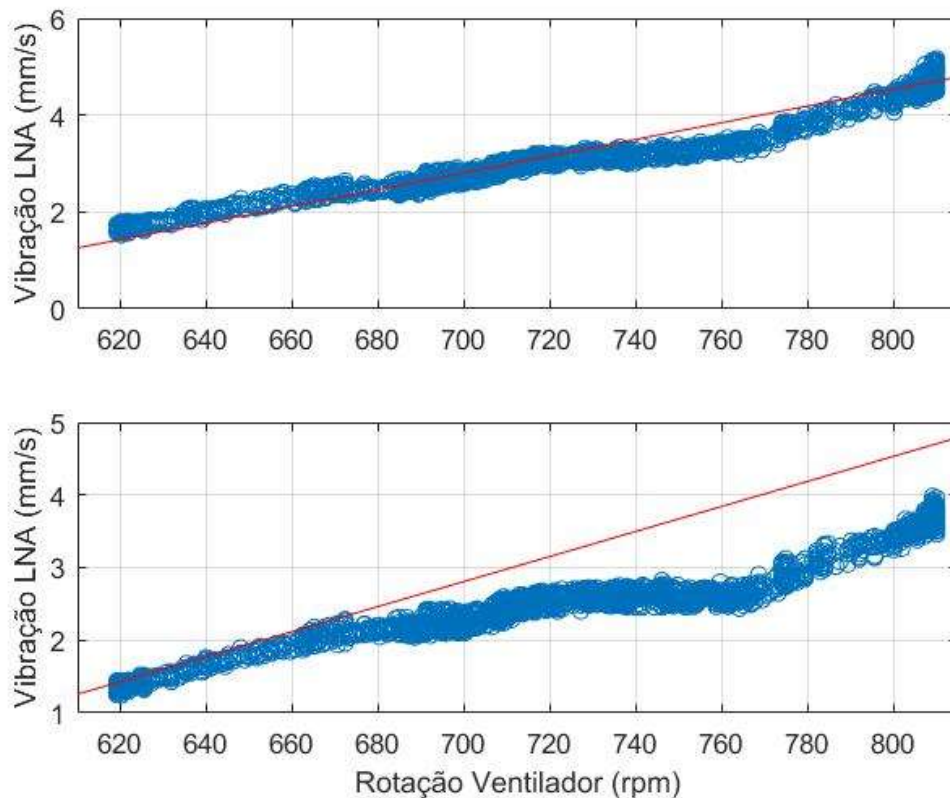


Figura 18 – Gráficos de dispersão rotação x vibrações

A seguir são apresentados os parâmetros utilizados em cada uma das metodologias a serem comparadas:

- AEML
  - Número de modelos: 2
  - Janela de dados q: 200 amostras
- AEML-R
  - Estrutura do modelo: MQO
  - Função de distância: norma euclidiana dos resíduos
  - Distância máxima permitida: 0,3
  - Quantidade mínima de amostras para geração de um modelo: 3% das amostras do conjunto de dados de treinamento
  - Número máximo de iterações 1.000
  - Número máximo de modelos: 2

- AEPOL
  - Ordem do polinômio: 3
- AEGRNN
  - Parâmetro de espalhamento: 0,2
- AERF
  - Quantidade máxima de árvores de decisão: 10
  - Profundidade das árvores: 50

Como descrito no capítulo 3, é utilizado o método de validação cruzada  $k$ -fold com  $k = 5$ . A Figura 19 mostra a norma total dos resíduos de Análise Externa de todas as seis metodologias para cada  $k$ -fold aplicados aos conjuntos de validação.

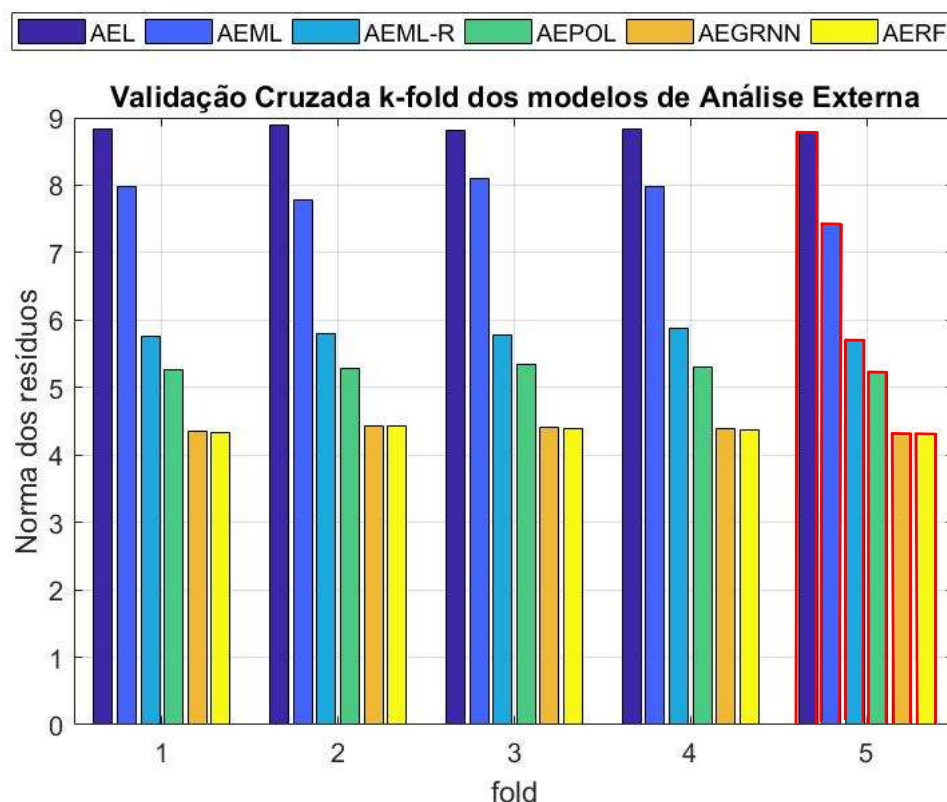


Figura 19 – Validação cruzada: Norma dos resíduos de Análise Externa aplicados aos conjuntos de dados de validação

Observa-se que as metodologias AERF e AGRNN possuem resultados semelhantes, com norma dos resíduos baixa em relação as outras metodologias. Assim como no estudo de caso apresentado na seção 4.1, a metodologia AEL obteve a maior norma dos resíduos, e para esta aplicação o método AEML apresentou resíduos elevados quando comparado com as outras metodologias.



Para este estudo de caso são escolhidos os modelos, associados aos conjuntos de dados da validação cruzada, que obtiverem as menores normas dos resíduos para cada uma das metodologias, sendo representados por um contorno vermelho nas barras da Figura 19.

Para validar se o classificador *Random Forest* está selecionando corretamente os modelos dos métodos AEML e AEML-R, vamos calcular a matriz de confusão, mostradas nas Figuras 20 e 21.

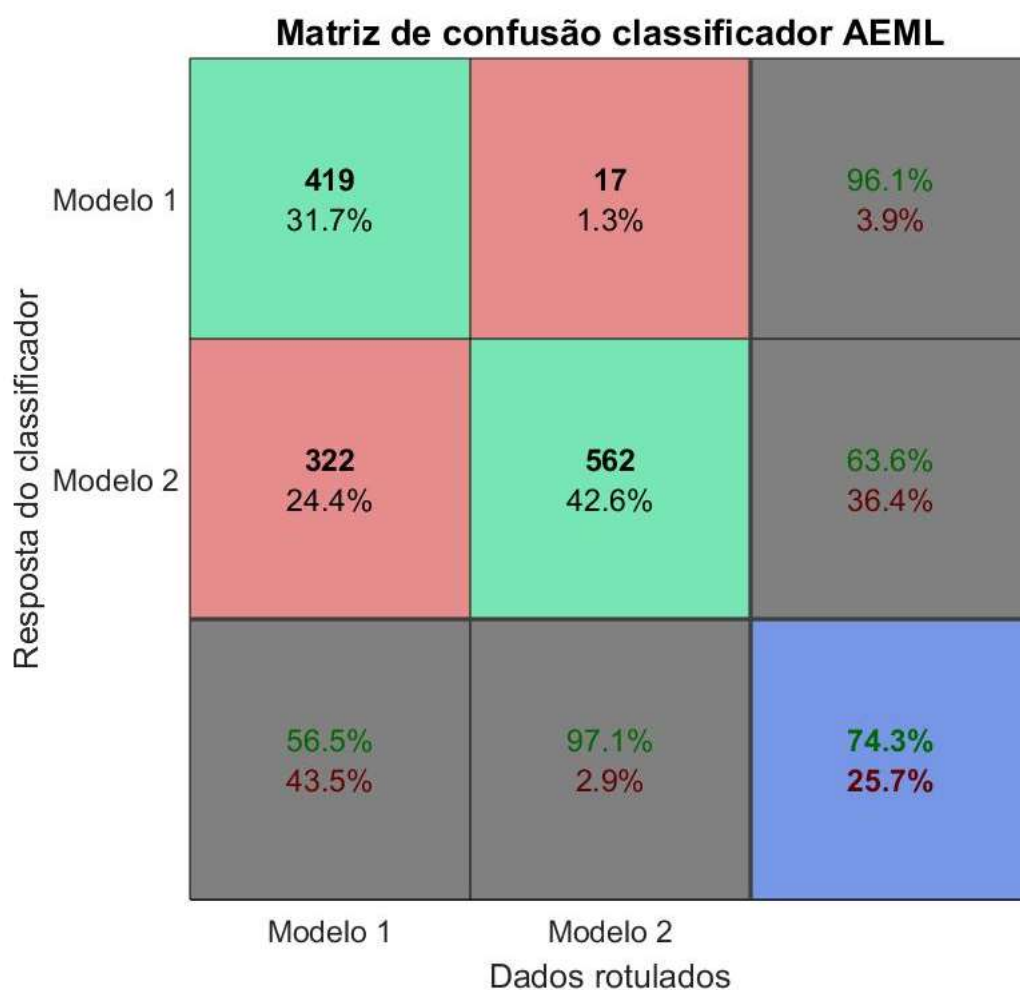


Figura 20 – Ventilador: Matriz de confusão classificador *Random Forest* AEML

Uma explicação detalhada da matriz de confusão foi feita no estudo de caso anterior. Observa-se na Figura 20 que o classificador conseguiu classificar 74,3% das amostras corretamente. Este resultado explica o motivo da norma dos resíduos do método AEML, mostrados na Figura 19, ser maior em relação aos métodos propostos neste trabalho, pois



em 25,7% das amostras o método está errando na seleção do modelo a ser usado no monitoramento de falhas.

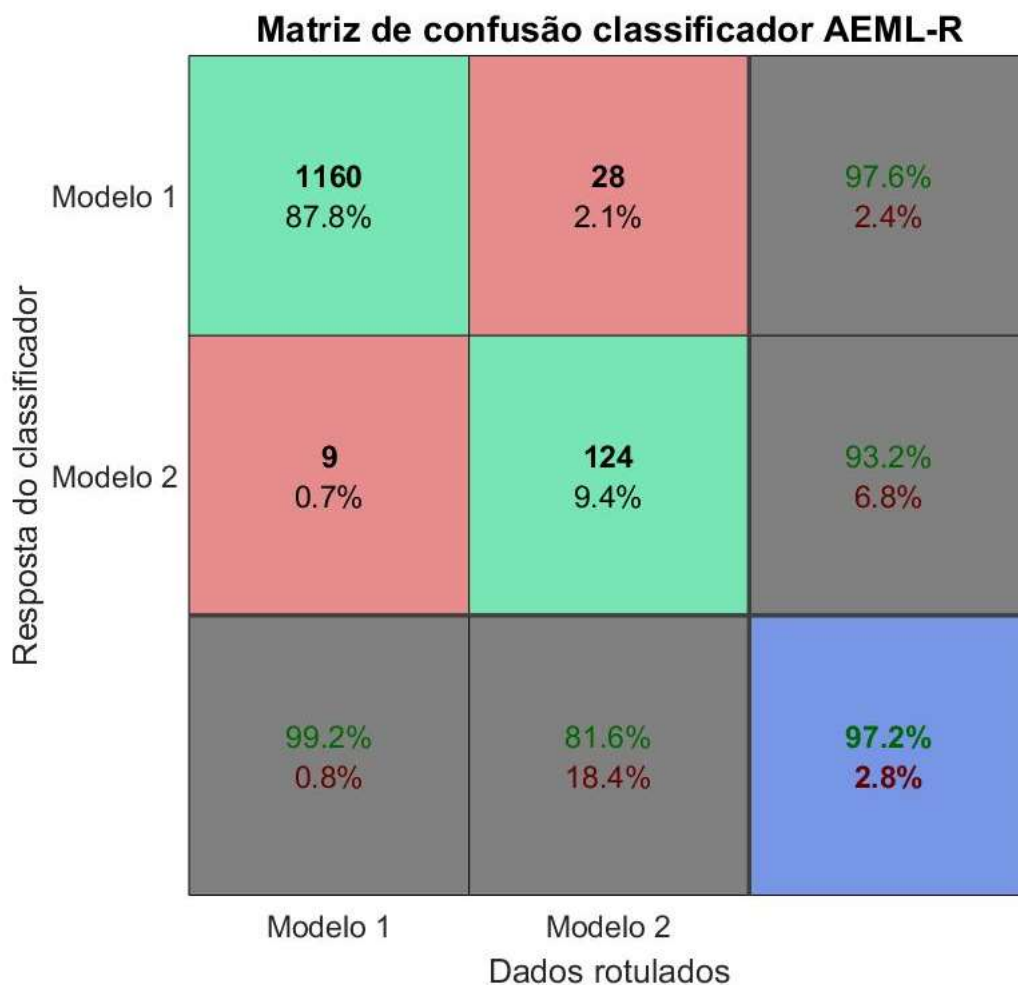


Figura 21 – Ventilador: Matriz de confusão classificador *Random Forest* AEML-R

Observa-se na Figura 21 que 97,2% das previsões estão corretas e 2,8% estão erradas.

A Figura 22 mostra a estatística  $T^2$  de Hotelling e os falsos positivos (FP) de todas as metodologias aplicados aos dados de validação.

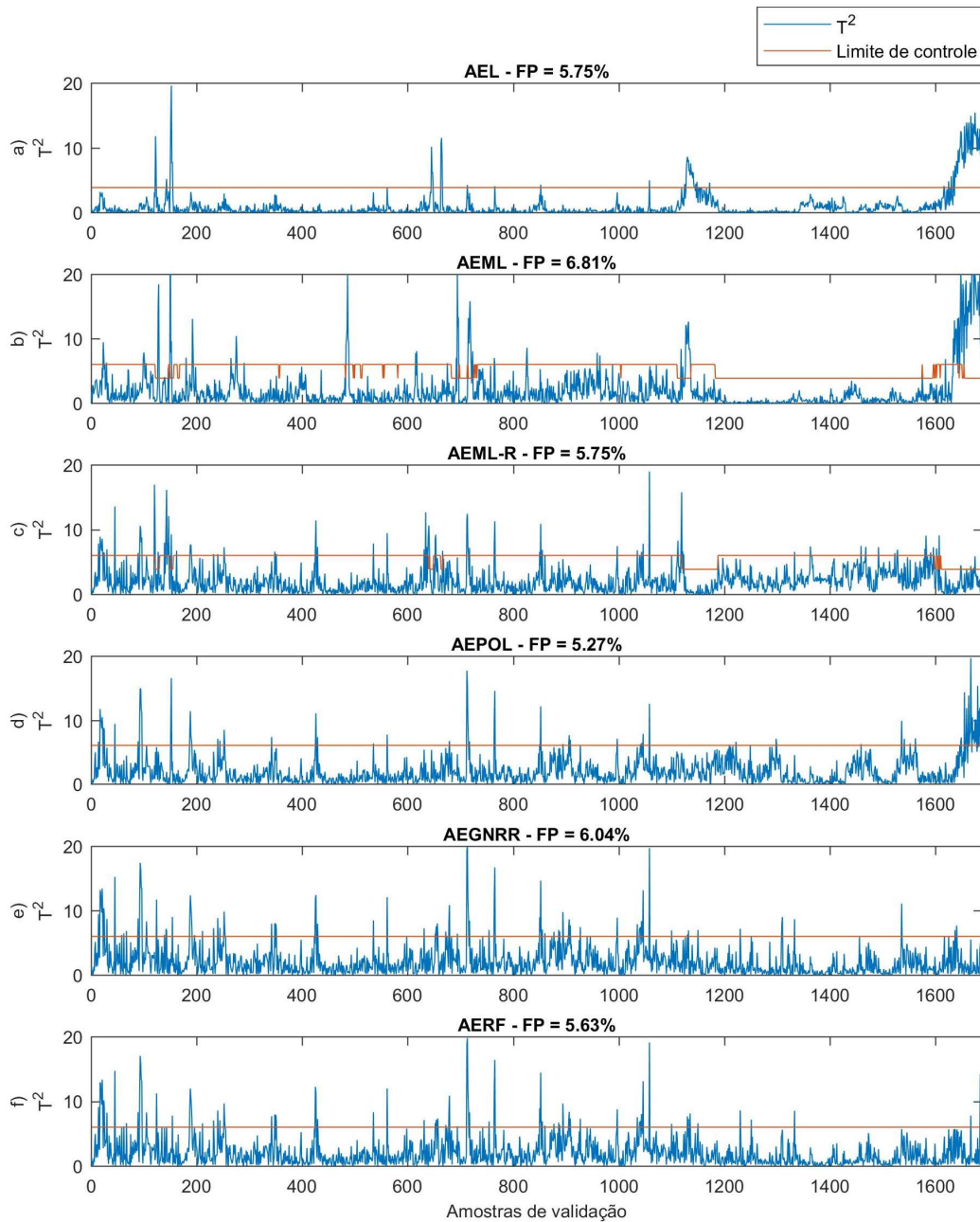


Figura 22 – Estatística  $T^2$  e avaliação de falsos positivos aplicados aos dados de validação. O gráfico a) mostra os resultados da metodologia AEL, b) AEML, c) AEML-R, d) AEPOL, e) AEGRNN e f) AERF

Como pode ser observado na Figura 22, para todas as metodologias os falsos positivos (alarmes falsos) possuem valores próximos ao valor esperado de 5%, tendo em vista que o limite de controle estatístico, em vermelho, foi calculado com o nível de confiança de 95%. Porém, a estatística  $T^2$  ao final do conjunto de amostras de validação ultrapassou o limiar estatístico para as metodologias AEL, AEML e AEPOL (em menor escala), indicando que estas metodologias não conseguiram capturar bem as relações entre as variáveis de rotação e vibração para as últimas amostras dos dados de validação.

Observa-se variações nos limiares estatísticos para as metodologias que utilizam múltiplos modelos (AEML e AEML-R). Isso ocorre, pois, cada modelo linear de Análise Externa está associado a um modelo PCA que possui um limiar específico, tendo em vista que a quantidade de amostras usadas no treinamento de cada modelo é diferente.

Para calcular os falsos negativos utilizaremos os dados mostrados na Figura 23. Por se tratar de um estudo de caso real, somente temos uma falha em um ponto de operação.

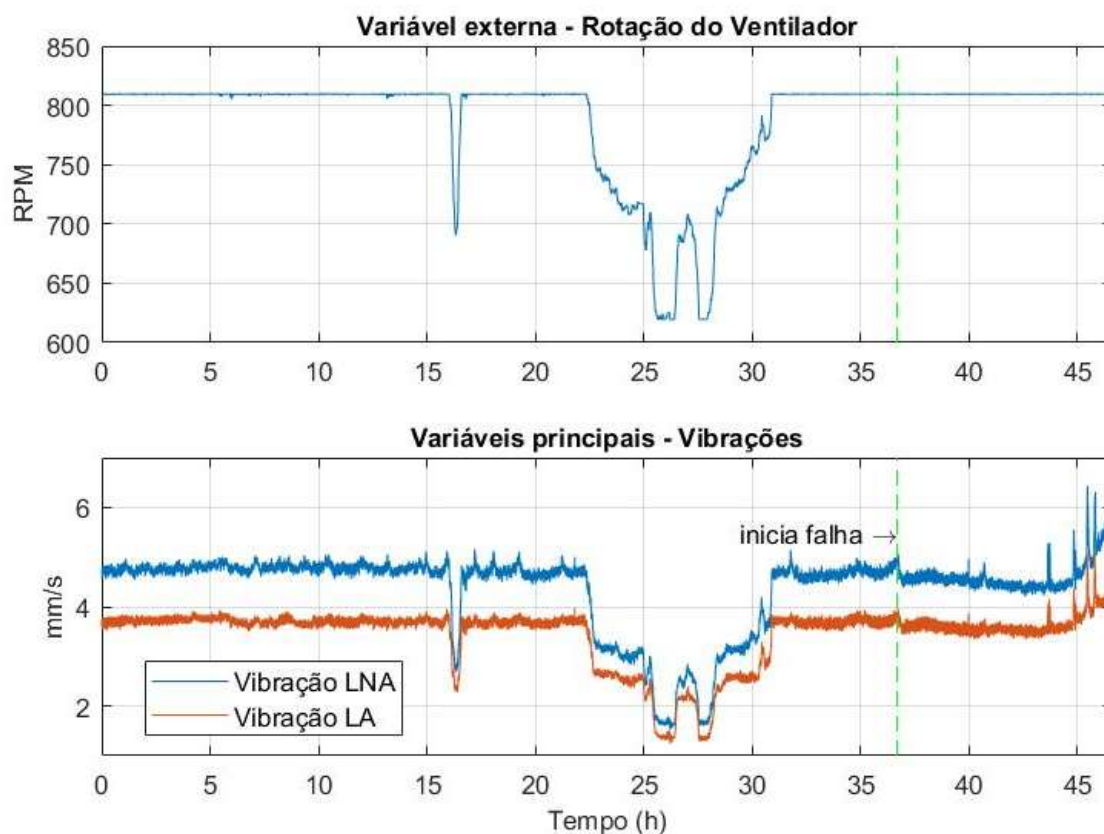


Figura 23 - Dados de rotação e vibração usados na avaliação de falsos negativos

Analisando a Figura 23, verifica-se que no momento de 36,71h (em verde) a vibração começa a variar sem que a rotação do ventilador se altere. Desta forma, é considerado que a falha se iniciou neste ponto.

A Figura 24 mostra a estatística  $T^2$  e avaliação de falsos negativos aplicados aos dados da Figura 23, assim como o gráfico da rotação do ventilador.

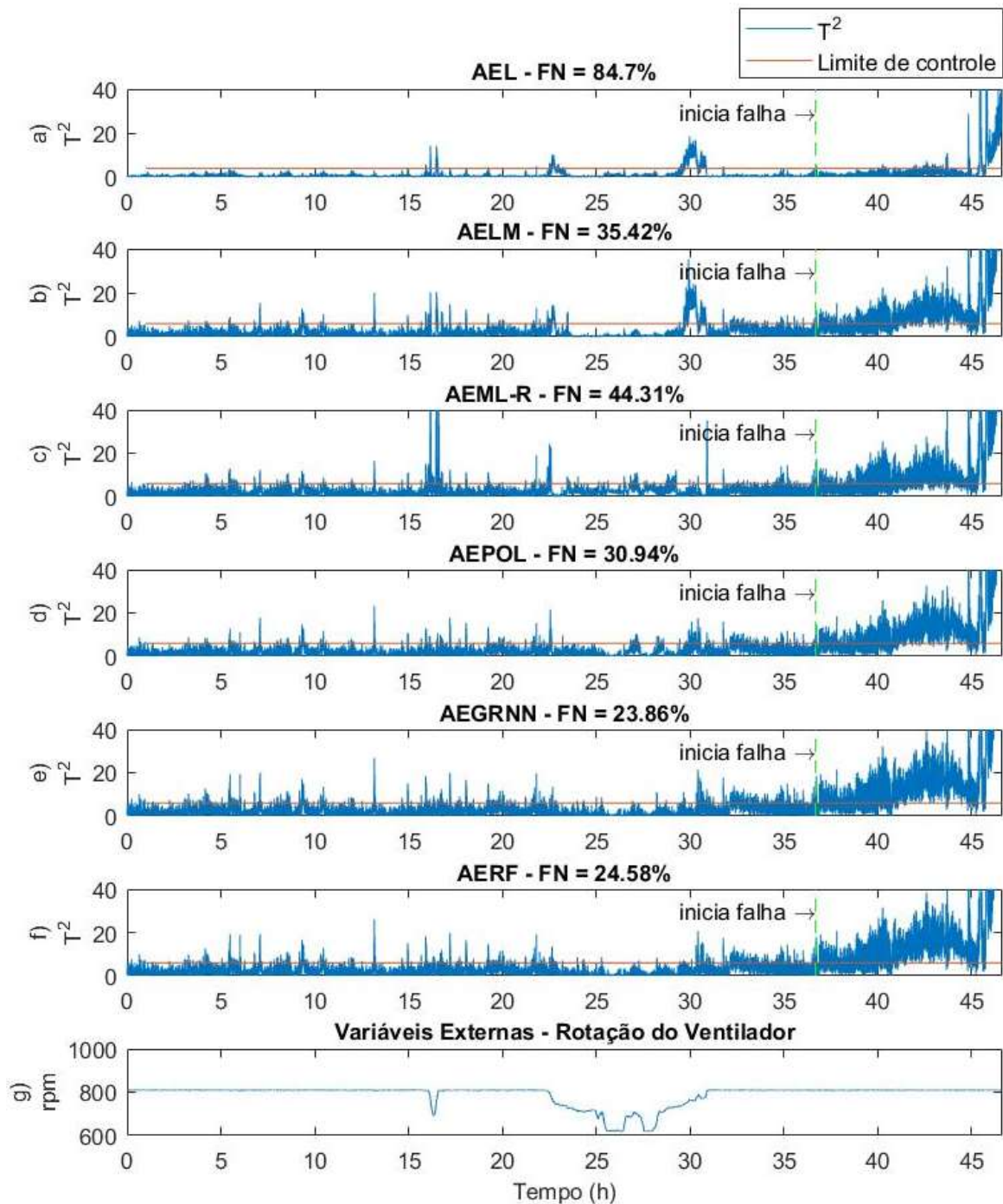


Figura 24 - Estatística  $T^2$  e avaliação de falsos negativos aplicados aos dados de falha. O gráfico a) mostra os resultados da metodologia AEL, b) AEML, c) AEML-R, d) AEPOL, e) AEGRNN, f) AERF e g) mostra as variações da rotação do ventilador

Observa-se que a metodologia AEL somente consegue detectar a falha após o instante de 45h, restando duas horas para que o ventilador parasse devido à quebra do casquilho e ponta do rotor. As outras metodologias conseguem detectar falhas com maior tempo de antecedência (estatística  $T^2$  acima do limiar), o que representa na prática maior tempo para tomada de ações corretivas que podem diminuir os danos causados aos componentes

e, consequentemente, diminuir o tempo de perda de produção devido a manutenção corretiva.

As metodologias AEGRNN (23,86%) e AERF (24,58%) obtiveram os melhores resultados, seguidos da AEPOL (30,94%), AEML (35,42%), AEML-R (44,31%) e AEL (84,7%).

## 5 Conclusão

Vários trabalhos vêm sendo desenvolvidos utilizando metodologias de controle estatísticos multivariado para detecção de falhas em processos industriais. Muitas vezes é suposto que as variáveis permanecem somente em um ponto de operação, o que limita o desenvolvimento para aplicações reais, tendo em vista que grande parte das variáveis de um processo são afetadas por fatores tais como mudanças de *setpoint* vindas de sistemas de otimização, adequação a demandas de produção, variações nas características dos insumos.

Neste trabalho são propostos cinco métodos com o objetivo de detectar falhas em processos industriais, que naturalmente possuem não linearidades, em múltiplos pontos de operação.

Os métodos são aplicados a dois estudos de caso, sendo uma simulação e um processo industrial real. Através dos índices de falsos positivos é observado que os alarmes falsos encontrados são os especificados para todos os métodos, com valores próximos ao nível de confiança usado para calcular os limiares estatísticos. Através dos índices de falsos negativos comprova-se que os métodos propostos foram capazes de detectar falhas, mesmo quando o método de Análise Externa linear proposto por Kano (2004) não foi.

Para a aplicação real de monitoramento de vibração de ventiladores, os métodos propostos neste trabalho possibilitaram a previsão da avaria do equipamento com cerca de dez horas de antecedência. Esta previsão pode gerar uma ação preventiva que pode evitar que os danos do equipamento rotativo se agravem, diminuindo assim as perdas de produção devido ao tempo de manutenção corretiva e consequente perda financeira da empresa.

Os métodos propostos funcionaram adequadamente, porém, os métodos que utilizam múltiplos modelos possuem parâmetros de complexo ajuste, dificultando o seu uso. Os métodos não lineares que utilizam GRNN e *Random Forest* são os mais indicados a serem utilizados, devido a facilidade de parametrização e bons resultados. O método AEML-R se destaca por possuir custo computacional elevado de treinamento, porém, na etapa *on line* de monitoramento de falhas, em que é crítico o custo computacional, o método possui processamento semelhante aos demais métodos.

As técnicas estudadas neste trabalho possuem pouca tolerância a erros de modelagem. Como em processos reais as relações entre as variáveis de processo não são constantes, é recomendado o estudo de técnicas de modelagem adaptativa, que possam

atualizar os modelos, sem perder a funcionalidade de distinguir as falhas de operações normais.

## 6 Referências Bibliográficas

ASSIS, A.; MUNARO, C.; CIARELLI, P. Monitoramento de processos industriais em diferentes regiões de operação utilizando análise externa não linear. **Conferência Brasileira de Dinâmica, Controle e Aplicações - DINCON 2017**, São José do Rio Preto-SP.

AL-DUWAISH, H.; NAEEM, Wasif. Nonlinear model predictive control of hammerstein and wiener models using genetic algorithms. In: **Control Applications, 2001.(CCA'01). Proceedings of the 2001 IEEE International Conference On**. IEEE, 2001. p. 465-469.

ARTHUR, David; VASSILVITSKII, Sergei. k-means++: The advantages of careful seeding. In: **Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms**. Society for Industrial and Applied Mathematics, 2007. p. 1027-1035.

BAKSHI, Bhavik R. Multiscale PCA with application to multivariate statistical process monitoring. **AIChE journal**, v. 44, n. 7, p. 1596-1610, 1998.

BREIMAN, Leo. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.

CHEN, Junhui; LIU, Kun-Chih. On-line batch process monitoring using dynamic PCA and dynamic PLS models. **Chemical Engineering Science**, v. 57, n. 1, p. 63-75, 2002.

CHERRY, Gregory A.; QIN, S. Joe. Multiblock principal component analysis based on a combined index for semiconductor fault detection and diagnosis. **IEEE Transactions on semiconductor manufacturing**, v. 19, n. 2, p. 159-172, 2006.

CHIANG, Leo H.; RUSSELL, Evan L.; BRAATZ, Richard D. **Fault detection and diagnosis in industrial systems**. Springer Science & Business Media, 2000.

CUI, Peiling; LI, Junhong; WANG, Guizeng. Improved kernel principal component analysis for fault detection. **Expert Systems with Applications**, v. 34, n. 2, p. 1210-1219, 2008.

DENG, Xiaogang; TIAN, Xuemin; CHEN, Sheng. Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault



diagnosis. **Chemometrics and Intelligent Laboratory Systems**, v. 127, p. 195-209, 2013.

DING, Mingtao; TIAN, Zheng; XU, Haixia. Adaptive kernel principal component analysis. **Signal Processing**, v. 90, n. 5, p. 1542-1553, 2010.

DUNIA, Ricardo et al. Identification of faulty sensors using principal component analysis. **AIChE Journal**, v. 42, n. 10, p. 2797-2812, 1996.

ELSHENAWY, Lamiaa M.; AWAD, Hamdi A. Recursive fault detection and isolation approaches of time-varying processes. **Industrial & Engineering Chemistry Research**, v. 51, n. 29, p. 9812-9824, 2012.

FARIA, J.; Munaro, C.; Ciarelli, P. Detecção de Falhas de Processos Não Lineares Via Análise Externa com Múltiplos Modelos. **XIII Simpósio Brasileiro de Automação Inteligente**, Porto Alegre-RS, pp. 2277-2282.

FINCH, Francis Eric. **Automated fault diagnosis of chemical process plants using model-based reasoning**. 1989. Tese de Doutorado. Massachusetts Institute of Technology.

FISCHLER, Martin A.; BOLLES, Robert C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: **Readings in computer vision**. 1987. p. 726-740.

FRIEDMAN, Jerome; HASTIE, Trevor; TIBSHIRANI, Robert. **The elements of statistical learning**. New York: Springer series in statistics, 2001.

GAO, Yong et al. Fault detection in time-varying chemical process through incremental principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 158, p. 102-116, 2016.

GE, Zhiqiang et al. Robust online monitoring for multimode processes based on nonlinear external analysis. **Industrial & Engineering Chemistry Research**, v. 47, n. 14, p. 4775-4783, 2008.

GE, Zhiqiang; YANG, Chunjie; SONG, Zhihuan. Improved kernel PCA-based monitoring approach for nonlinear processes. **Chemical Engineering Science**, v. 64, n. 9, p. 2245-2255, 2009.

GU, Shanmao et al. Fault Detection Approach Based on Weighted Principal Component Analysis Applied to Continuous Stirred Tank Reactor. **Open Mechanical Engineering Journal**, v. 9, p. 966-972, 2015.

HARROU, Fouzi et al. PLS-based EWMA fault detection strategy for process monitoring. **Journal of Loss Prevention in the Process Industries**, v. 36, p. 108-119, 2015.

HOSSIN, M.; SULAIMAN, M. N. A review on evaluation metrics for data classification evaluations. **International Journal of Data Mining & Knowledge Management Process**, v. 5, n. 2, p. 1, 2015.

HUANG, Jian; YAN, Xuefeng. Dynamic process fault detection and diagnosis based on dynamic principal component analysis, dynamic independent component analysis and Bayesian inference. **Chemometrics and Intelligent Laboratory Systems**, v. 148, p. 115-127, 2015.

JIN, Hyung Dae et al. Robust recursive principal component analysis modeling for adaptive monitoring. **Industrial & engineering chemistry research**, v. 45, n. 2, p. 696-703, 2006.

KANO, Manabu et al. Evolution of multivariate statistical process control: application of independent component analysis and external analysis. **Computers & chemical engineering**, v. 28, n. 6-7, p. 1157-1166, 2004.

KOMULAINEN, Tiina; SOURANDER, Mauri; JÄMSÄ-JOUNELA, Sirkka-Liisa. An online application of dynamic PLS to a dearomatization process. **Computers & Chemical Engineering**, v. 28, n. 12, p. 2611-2619, 2004.

KHEDIRI, Issam Ben; LIMAM, Mohamed; WEIHS, Claus. Variable window adaptive kernel principal component analysis for nonlinear nonstationary process monitoring. **Computers & Industrial Engineering**, v. 61, n. 3, p. 437-446, 2011.

KOMULAINEN, Tiina; SOURANDER, Mauri; JÄMSÄ-JOUNELA, Sirkka-Liisa. An online application of dynamic PLS to a dearomatization process. **Computers & Chemical Engineering**, v. 28, n. 12, p. 2611-2619, 2004.

LEE, Gibaek; HAN, Chonghun; YOON, En Sup. Multiple-fault diagnosis of the Tennessee Eastman process based on system decomposition and dynamic PLS. **Industrial & engineering chemistry research**, v. 43, n. 25, p. 8037-8048, 2004.

LEE, Hae Woo; LEE, Min Woo; PARK, Jong Moon. Multi-scale extension of PLS algorithm for advanced on-line process monitoring. **Chemometrics and Intelligent Laboratory Systems**, v. 98, n. 2, p. 201-212, 2009.

LI, Weihua et al. Recursive PCA for adaptive process monitoring. **Journal of process control**, v. 10, n. 5, p. 471-486, 2000.

LIU, Xueqin et al. Moving window kernel PCA for adaptive monitoring of nonlinear processes. **Chemometrics and intelligent laboratory systems**, v. 96, n. 2, p. 132-143, 2009.

MU, Shengjing et al. Online dual updating with recursive PLS model and its application in predicting crystal size of purified terephthalic acid (PTA) process. **Journal of Process Control**, v. 16, n. 6, p. 557-566, 2006.

MURADORE, Riccardo; FIORINI, Paolo. A PLS-based statistical approach for fault detection and isolation of robotic manipulators. **IEEE Transactions on Industrial Electronics**, v. 59, n. 8, p. 3167-3175, 2012.

OYELEYE, Olayiwola Oluwemimo. **Qualitative modeling of continuous chemical processes and applications to fault diagnosis**. 1989. Tese de Doutorado. Massachusetts Institute of Technology.

PATAN, Krzysztof. **Artificial neural networks for the modelling and fault diagnosis of technical processes**. Springer, 2008.

PORTNOY, Ivan et al. An improved weighted recursive PCA algorithm for adaptive fault detection. **Control Engineering Practice**, v. 50, p. 69-83, 2016.

QIN, S. Joe. Recursive PLS algorithms for adaptive data modeling. **Computers & Chemical Engineering**, v. 22, n. 4-5, p. 503-514, 1998.

QIN, S. Joe. Survey on data-driven industrial process monitoring and diagnosis. **Annual reviews in control**, v. 36, n. 2, p. 220-234, 2012.

QIN, S. Joe; ZHENG, Yingying. Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures. **AIChE Journal**, v. 59, n. 2, p. 496-504, 2013.

RATO, Tiago J.; REIS, Marco S. Fault detection in the Tennessee Eastman benchmark process using dynamic principal components analysis based on decorrelated residuals (DPCA-DR). **Chemometrics and Intelligent Laboratory Systems**, v. 125, p. 101-108, 2013.

RONGYU, L. I.; GANG, R. O. N. G. Fault Isolation by Partial Dynamic Principal Component Analysis in Dynamic Process1. **Chinese Journal of Chemical Engineering**, v. 14, n. 4, p. 486-493, 2006.

SHAMS, MA Bin; BUDMAN, H. M.; DUEVER, T. A. Fault detection, identification and diagnosis using CUSUM based PCA. **Chemical Engineering Science**, v. 66, n. 20, p. 4488-4498, 2011.

SPECHT, Donald F. A general regression neural network. **IEEE transactions on neural networks**, v. 2, n. 6, p. 568-576, 1991.

VENKATASUBRAMANIAN, Venkat et al. A review of process fault detection and diagnosis: Part III: Process history based methods. **Computers & chemical engineering**, v. 27, n. 3, p. 327-346, 2003.

TAN, Shuai et al. Multimode process monitoring based on mode identification. **Industrial & Engineering Chemistry Research**, v. 51, n. 1, p. 374-388, 2011.

TONG, Chudong; PALAZOGLU, Ahmet; YAN, Xuefeng. An adaptive multimode process monitoring strategy based on mode clustering and mode unfolding. **Journal of Process Control**, v. 23, n. 10, p. 1497-1507, 2013.

TORR, Philip HS; ZISSERMAN, Andrew. MLESAC: A new robust estimator with application to estimating image geometry. **Computer vision and image understanding**, v. 78, n. 1, p. 138-156, 2000.

YIN, Shen; ZHU, Xiangping; KAYNAK, Okyay. Improved PLS focused on key-performance-indicator-related fault diagnosis. **IEEE Transactions on Industrial Electronics**, v. 62, n. 3, p. 1651-1658, 2015.

YONG, Gao; XIN, Wang; ZHENLEI, Wang. Fault detection for a class of industrial processes based on recursive multiple models. **Neurocomputing**, v. 169, p. 430-438, 2015.

WANG, Xun; KRUGER, Uwe; LENNOX, Barry. Recursive partial least squares algorithms for monitoring complex industrial processes. **Control Engineering Practice**, v. 11, n. 6, p. 613-632, 2003.

YU, Jie; QIN, S. Joe. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. **AIChE Journal**, v. 54, n. 7, p. 1811-1829, 2008.

ZHAO, Shi Jian; ZHANG, Jie; XU, Yong Mao. Monitoring of processes with multiple operating modes through multiple principle component analysis models. **Industrial & engineering chemistry research**, v. 43, n. 22, p. 7025-7035, 2004.

ZHANG, Yingwei et al. Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares. **IEEE Transactions on Industrial Informatics**, v. 6, n. 1, p. 3-10, 2010.

ZHANG, Yingwei; LI, Shuai; TENG, Yongdong. Dynamic processes monitoring using recursive kernel principal component analysis. **Chemical Engineering Science**, v. 72, p. 78-86, 2012.

ZHANG, Yingwei; LI, Shuai; HU, Zhiyong. Improved multi-scale kernel principal component analysis and its application for fault detection. **Chemical Engineering Research and Design**, v. 90, n. 9, p. 1271-1280, 2012.

ZHENG, Ying; QIN, S. Joe; CHAI, Tianyou. Fault detection of multimode processes using concurrent projection to latent structures. **IFAC-PapersOnLine**, v. 49, n. 7, p. 705-710, 2016.